

THE EFFICACY OF PROFILE MATCHING AS A MEANS OF CONTROLLING
FOR THE EFFECTS OF RESPONSE DISTORTION ON PERSONALITY
MEASURES

A Dissertation

by

RYAN MATTHEW GLAZE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Winfred Arthur, Jr.
Committee Members,	David Martin
	Stephanie C. Payne
	John Edens
Head of Department,	Paul Wellman

December 2012

Major Subject: Psychology

Copyright 2012 Ryan Matthew Glaze

ABSTRACT

Researchers and practitioners continue to be concerned about the magnitude, extent, and effects of response distortion when self-report personality measures are used in high-stakes testing. One method for mitigating response distortion that has not received much empirical attention is profile matching. Profile matching assesses the fit between test-takers' predictor profiles and a standard profile which represents an ideal or high performing employee's profile. Since profile matching assesses fit, it can capture nonlinear effects. Furthermore, high predictor scores are not necessarily associated with higher criterion scores. Test-takers who distort their responses by choosing inaccurately extreme response options may improve their chances of being hired if a linear model is used, but this approach is unlikely to be effective if a profile matching strategy is used as long as the standard profile is unknown to the test-takers. As such, the primary objective of the present study was to examine the extent to which profile matching may alleviate concerns about response distortion. A secondary objective was to examine characteristics of the standard profile that are associated with the efficacy of this approach.

The present study compared the effects of response distortion on personality test scores, and their criterion-related validity in predicting tenure, based on a linear composite and a profile fit score. The present study used data from 996 applicants who completed a personality test in a high-stakes testing context. Missing data were imputed for a subset of applicants who did not complete two response distortion scales. As such,

the results provided an initial proof-of-concept of the effectiveness of profile matching as a personnel decision-making strategy using a blend of real and simulated data. The results suggest that profile fit scores are less related to response distortion and display higher criterion-related validity than linear composite scores. However, the difference in criterion-related validity could not be attributed to response distortion. The results further suggest that the amount of scatter in the standard profile is negatively associated with the profile fit score's susceptibility to response distortion and positively related to criterion-related validity.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Winfred Arthur, Jr., and my committee members, Dr. Stephanie Payne, Dr. John Edens, and Dr. David Martin, for their guidance, support, and insight throughout the course of this research. Additionally, I would like to thank Dr. Aaron Taylor for his thoughtful reviews and Dr. Jason Taylor for helping make this project possible.

Thanks also go to my friends and colleagues at Texas A&M University. Special thanks go to my friends and colleagues Eswen Fava, Steven Jarrett, Ira Schurig, Gonzalo Muñoz–Galvez, Jennifer McDonald, and Andrew Naber for their support and friendship.

Finally, thanks go to my mom, dad, and brother for their encouragement, support, patience, and love.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
 CHAPTER	
I INTRODUCTION.....	1
II RESPONSE DISTORTION.....	6
Methods for Investigating Response Distortion.....	8
Lie Scales	8
Contrasting Groups	17
Response Distortion Debate	20
Item Exposure and Coaching Threats	23
Effects of Response Distortion.....	27
Construct-Related Validity	27
Criterion-Related Validity	28
Techniques to Detect and Deter Response Distortion.....	30
Warnings	31
Forced-Choice Response Formats	32
III PROFILE MATCHING	35
Nonlinear Effects.....	39
Personality Variables and Nonlinear Relationships	45
Whole Person Approaches	47
IV CURRENT STUDY	55
Susceptibility to Response Distortion	55
Criterion-Related Validity	58

	Subgroup Differences.....	61
	Profile Configuration.....	63
	Assessing Fit	70
V	METHOD.....	72
	Participants.....	72
	Measures.....	72
	Personality	72
	Improbable Item Lie Scale	73
	Consistency–In–Responding Lie Scale	74
	Tenure.....	74
	Profile Fit Scores	75
	Simulated Profile Fit Scores.....	75
	Linear Composite Scores	77
	Procedure.....	78
VI	RESULTS.....	79
	Empirical Profile Fit Scores	79
	Relationship Between Profile Fit and Linear	
	Composite Scores.....	79
	Susceptibility to Response Distortion	79
	Criterion–Related Validity	80
	Subgroup Differences.....	83
	Profile Scatter	85
	Susceptibility to Response Distortion	85
	Criterion–Related Validity	87
VII	DISCUSSION AND CONCLUSION	95
	Critiques of Profile Matching.....	101
	Limitations and Directions for Future Research	103
	Practical and Scientific Implications.....	108
	REFERENCES	111

LIST OF FIGURES

FIGURE	Page
1 An Illustrative Example of 3 Score Profiles.....	39
2 An Illustrative Example of an Asymptotic Relationship	40
3 An Illustrative Example of an Inverted–U Relationship	41
4 An Illustrative Example of a Cubic Relationship.....	43
5 An Illustrative Example of a Multiple Variable Asymptotic Relationship	44
6 An Illustrative Example of a Circumplex for Extraversion (I) and Agreeableness (II)	52
7 Three Example Profiles which Differ in Terms of Profile Complexity	64
8 Standard Profiles for Voluntary Turnover and Tenure	105

LIST OF TABLES

TABLE		Page
1	Illustrative Profile Component Scores, Elevation, and Scatter	66
2	Means, Standard Deviations, and Sample Sizes for the Two Referent Groups and the Overall Sample.....	76
3	Standard Profiles for Voluntary Turnover and Tenure	77
4	Simulated Standard Profiles	78
5	Means, Standard Deviations, and Correlations for Study Variables.....	84
6	Subgroup Differences for Profile Fit and Linear Composite Scores	85
7	Summary of Research Hypotheses and Results	91
8	Intercorrelations Amongst FFM Variables Reported by Ones (1993) and From the Current Study	104

CHAPTER I

INTRODUCTION

The use of self-report noncognitive tests in general and personality tests in particular for personnel decision-making is quite widespread. Concomitant with the use of and research on these classes of tests are concerns regarding the extent, magnitude, and effects of response distortion when said tests are used in high-stakes testing contexts (e.g., Arthur, Glaze, Villado, & Taylor, 2010; Cullen, Sackett, & Lievens, 2006; Landers, Sackett & Tuzinski, 2011; Levashina, & Campion, 2007; Schmitt & Kuncce, 2002). Although there appears to be a growing consensus that most, if not all, noncognitive self-report tests are susceptible to response distortion to some extent, there is continued debate regarding the extent and magnitude of response distortion that occurs in applicant settings and the effects of response distortion. Concerns regarding response distortion are exacerbated in large-scale testing efforts and where unproctored internet-based tests are used, as item exposure and availability in the public domain becomes a major concern in these situations. Although several techniques to detect and deter response distortion have been developed, these techniques are unlikely to be effective if test-takers are aware of their presence.

Profile matching is one potential response distortion deterrence technique that has not received much empirical attention. Profile matching entails assessing the fit between a standard predictor profile and a test-taker's predictor profile. A predictor profile is defined as the arrangement of test-takers' scores. As such, profile matching strategies capture both the elevation and scatter of profile scores. Elevation refers to the

mean of a test-taker's scores. Scatter refers to the configuration of a test-taker's scores (i.e., the extent to which each of the test-takers' predictor scores deviates from that test-taker's elevation).

Historically, clinical and counseling psychologists have relied on both the elevation and scatter of diagnostic test scores in diagnosing and classifying patients. Profile matching is one method in which clinicians diagnose patients and determine the extent to which the patient represents a specified diagnostic category (Greene, 2000). That is, they compare the patient's score profile to normative profiles from other individuals in the same category to determine which diagnostic profile best matches the patient's profile. Then, clinicians will typically assess how similar the patient's profile is to the prototypical profile of that diagnostic category (e.g., Glenn, Beckham, Sampson, Feldman, Hertzberg, & Moore, 2002; Sheppard, Smith, & Rosenbaum, 1988). Several diagnostic inventories rely on profile types to classify patients, including the Personality Inventory for Children (Wirt, Lachar, Klinedinst, & Seat, 1984) and the Minnesota Multiphasic Personality Inventory (Dahlstrom, Welsh, & Dahlstrom, 1972).

Since profile matching is a decision-making strategy that assesses the fit between an ideal or standard profile and a test-taker's scores, it does not rely on a linear model. However, the preponderance of personality research in industrial and organizational (I/O) psychology has relied on linear models, although there is some evidence that suggests the relationship between personality and performance may be nonlinear (e.g., Le, Oh, Robbins, Ilies, Holland, & Westrick, 2010). The extant literature provides conflicting conclusions regarding whether linear or nonlinear models are

superior in terms of criterion-related validity (e.g., Le et al., 2011; Robie & Ryan, 1999). It is important to note that the purpose of the current study is to investigate the extent to which profile matching may alleviate concerns regarding response distortion. Profile matching may differ from linear models in the extent to which it is susceptible to response distortion and its criterion-related validity. That is, profile fit scores differ from linear composite scores in terms of the proposed relationship between these scores and performance, and also potentially differ in terms of their susceptibility to response distortion. The viability of profile matching is contingent upon the use of profile fit scores resulting in lower susceptibility to response distortion compared to scores based on a linear model, and criterion-related validity that is equal to or greater than that of scores based on a linear model. That is, it is unlikely that researchers and practitioners would be willing to use a technique that reduces the effects of response distortion while also reducing criterion-related validity. However, since response distortion is commonly conceptualized as construct-irrelevant variance, reducing response distortion is likely to result in increased criterion-related validity. In order to examine their relative susceptibility to response distortion, both the profile fit and linear composite scores were corrected for response distortion with the expectation that the criterion-related validity of profile fit scores would be less affected by said correction compared to the linear composite scores.

The primary objective of the current study was to investigate and document the extent to which concerns about response distortion may be alleviated by use of profile matching strategies in personnel selection decision-making. Profile fit scores were

compared to linear composite scores in terms of their susceptibility to response distortion. It was posited that profile fit scores would be less susceptible to response distortion. As such, it was expected that profile fit scores and linear composite scores would display different patterns of relationships with various lie scale scores and turnover (i.e., voluntary turnover and tenure). Additionally, since linear composite scores and profile fit scores were posited to represent different manifestations of the personality domain space, and personality profile fit scores have not received much empirical attention in the extent literature, the current study also sought to examine and document potential subgroup differences in personality profile fit scores.

A secondary objective was to examine the extent to which characteristics of standard profiles influence the efficacy of profile matching in alleviating concerns regarding response distortion. Specifically, standard profiles were generated that differed in their configuration. These various standard profiles were then compared in terms of their susceptibility to response distortion, relationship with turnover, and patterns of subgroup differences.

To achieve the study objectives, the data for the current study must display relationships between the personality variables and turnover. Furthermore, there must be personality, turnover, and response distortion data for each test-taker. The dataset did not meet all of these boundary conditions. Specifically, a proportion of the test-takers did not complete response distortion measures. Response distortion data were imputed for these test-takers on the basis of their personality test scores. Given the use of a dataset that contains both real and simulated data, the present study represents a proof-

of-concept. As such, the results of the current study should be interpreted in a similar manner as results from simulation studies. Furthermore, like simulation studies, it is critical that future field tests of this proof-of-concept be conducted to determine the veracity and replicability of the findings of the current study.

CHAPTER II

RESPONSE DISTORTION

In personnel selection contexts, response distortion may best be conceptualized as a specific form of malfeasant behavior on noncognitive measures and it entails deliberately falsifying one's responses to test items in an effort to create an overly positive impression (Arthur & Glaze, 2011). Several terms and descriptive labels have been used to describe this type of behavior including faking, socially desirable responding, self-enhancement, and lying. Paulhaus (2002) distinguishes between impression management and self-deceptive enhancement as two facets of socially desirable responding. Impression management entails intentionally manipulating one's responses in order to produce a favorable impression. It involves purposefully exaggerating socially desirable characteristics and minimizing or denying socially undesirable characteristics. Self-deceptive enhancement refers to unconsciously viewing oneself in an inaccurately positive manner and typically requires a lack of personal insight and self-awareness. Researchers tend to focus on impression management, as situational factors (e.g., high- versus low-stakes testing contexts) have little to no bearing on self-deceptive enhancement.

These conceptualizations of response distortion (e.g., malfeasant behavior, faking) typically consider response distortion as a form of systematic error variance which contaminates personality test scores (Arthur, Woehr, & Graziano, 2001). In the absence of techniques to deter response distortion, test-takers may distort their responses when the most desirable response is apparent to them. Response distortion threatens the

efficacy of noncognitive measures to the extent that test-takers differentially engage in response distortion, such that said response distortion influences the rank order of test-taker scores. Despite the minority claims that response distortion is constant over test-takers (e.g., Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt 2007a, 2007b), the vast majority of research suggests that test-takers vary in the extent to which they engage in response distortion (e.g., Arthur et al., 2010; Griffith, Chmielowski, & Yoshita, 2007; McFarland & Ryan, 2000).

Researchers have proposed alternate conceptualizations of response distortion. For example, some researchers have argued that response distortion may be better conceptualized as a substantive personality trait (Block, 1965; Costa & McCrae, 1988; Nicholson & Hogan, 1990) which may display positive correlations with job performance in occupations requiring interpersonal interactions (cf. Viswesvaran, Ones, & Hough, 2001; Ones, Viswesvaran, & Reiss, 1996) and predict other outcomes as well (e.g., subjective well-being; Kozma & Stones, 1987; McCrae, 1986). Hogan's (1998) socio-analytic theory posits that individuals negotiate and maintain their personal reputation through interpersonal interactions, and completing personality tests is an interpersonal interaction between the test-taker and an (unknown) other. Thus, socially desirable responding may reflect differences in effective socialization, and individuals who respond in a socially desirable manner on personality tests are likely to respond similarly in other social interactions. However, meta-analytic results suggest that the relationship between response distortion (as measured by lie scales) and managerial

performance is quite low ($r = .04$, $SD_r = 0.03$, $k = 17$, $N = 20,069$; Viswesvaran et al., 2001; see also Ones et al., 1996; Li & Bagger, 2006; White, Young, & Rumsey, 2001).

Although alternative conceptualizations of response distortion exist, researchers who conceptualize response distortion as systematic error variance are concerned with the extent to which selection tools are susceptible to response distortion. As such, researchers have investigated response distortion within the context of numerous selection devices including situational judgment tests (Cullen et al., 2006; Peeters & Lievens, 2005), biodata (Kluger, & Collela, 1993; Levashina, Morgeson, & Campion, 2009; Schmitt & Kuncze, 2002), employment interviews (Delery, & Kacmar, 1998; Ellis, West, Ryan, & DeShon, 2002; Levashina, & Campion, 2007), and assessment centers (McFarland, Yen, Harold, Viera, & Moore, 2005) with the majority of research in this domain investigating self-reported personality (e.g., Arthur et al., 2010; Landers et al., 2011).

Methods for Investigating Response Distortion

The results of research investigating the extent of response distortion typically vary as a function of the methods used (e.g., McFarland & Ryan, 2006). Several methods for investigating the extent of response distortion have been developed, including the use of lie scales and contrasting groups (e.g., applicants versus incumbents, instructed-faking versus instructed-honesty).

Lie Scales. Lie scales broadly refer to a class of measures that attempt to assess the tendency to engage in response distortion. Lie scales typically are composed of items that are socially desirable and unlikely to be true (i.e., improbable items). Lie

scales can take one of two forms. External lie scales are independent measures developed to assess response distortion. Examples of external lie scales are the Unlikely Virtues Scale (Hough, 1998), and the Marlowe–Crowne Social Desirability Scale (Crowne & Marlowe, 1960). Embedded lie scales involve the insertion of items designed to detect response distortion into the focal measure. Examples of personality scales that use embedded lies scales include the California Psychological Inventory (Gough & Bradely, 1996), the Hogan Personality Inventory (Hogan & Hogan, 1992), the Occupational Personality Questionnaire (SHL Group, 2000), the Personality Research Form (Jackson, 1999), and the Inwald Personality Inventory (Inwald, 1992). Lie scales can be used to correct for response distortion by adjusting the focal scale scores in a manner commensurate with the amount of suspected distortion. Furthermore, if a predetermined lie scale score is exceeded, the test-taker’s score can be flagged and invalidated.

Although the aforementioned lie scales are frequently used in research and practice (Goffin & Christiansen, 2003), there are several concerns regarding their use in applied settings. Studies that have examined statistical control techniques have found that these techniques reduce the effects of response distortion, but have little to no effect on the criterion-related validities of personality test scores (e.g., Hough, 1998; Ones, et al., 1996). Furthermore, as Cronbach (1990) wrote “Once test users take a wrong course, there is no going back to that choice point” (p. 521). Thus, score adjustments do not necessarily reflect the test-taker’s true standing on the construct of interest.

The use of implausible items is vulnerable to false positive results (McCrae & Costa, 1983). For example, a test-taker who is extremely conscientious may have a high score on a lie scale even though no response distortion occurs. The effectiveness of lie scales has been questioned (McFarland & Ryan, 2006), and research suggests that test-takers may be coached to inflate their test scores on the focal personality measures without detection via the lie scale (Alliger, Lilienfeld, & Mitchell, 1996; Dwight & Alliger, 1997; Kroger & Turnbull, 1975). This may be due to the fact that most lie scales were not developed for occupational settings. For example, the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991) contains items that are inappropriate for selection contexts (i.e., invasive items). Furthermore, the MMPI lie scales (i.e., L- and K-scales) were initially developed as a diagnostic tool for assessing mental health. As such, the use of these scales is not appropriate for selection contexts due to the American's with Disabilities Act (ADA). Indeed, Paulhus (1986) argues that lie scales generally have poor construct validity (see also Griffith & Peterson, 2008; McFarland & Ryan, 2006). Most lie scales tend to have poor convergent validity (Holden & Fekken, 1989) and poor factor structures (Holden & Fekken, 1989; Paulhus, 1991). Lie scales have also demonstrated meaningful relationships with emotional stability and conscientiousness (Ones et al., 1996). These concerns have led several researchers to question the efficacy and appropriateness of relying on lie scales to detect response distortion (e.g., Griffith & Peterson, 2008; McFarland & Ryan, 2006; Morgeson et al., 2007a, 2007b), and calls for additional methods of detecting response distortion have been made (e.g., Oswald & Hough, 2008; Kuncel & Borneman, 2007).

Although the improbable item approach appears to be quite common, there are several other approaches to constructing lie scales. The use of bogus items (Anderson, Warner, & Spencer, 1984) entails asking test-takers to rate their experience with a number of tasks, where a small number of these tasks are imaginary (e.g., “matrixing solvency files”). Anderson et al. reported that approximately 45% of applicants indicated having experience with these bogus tasks. Another approach is to measure the consistency with which test-takers respond to items. Consistency-in-responding approaches can take a variety of forms. For example, Hand (1964) developed the Acquiescence scale by selecting pairs of items from the Guilford-Zimmerman Temperament Survey (Guilford, Zimmerman, & Guilford, 1976) where the items measured the same trait, were similar in social desirability levels, and the valences were opposing (i.e., the keyed response for one item was “yes” and the other was “no”). This approach resulted in a 144-item scale which was designed to measure acquiescence (i.e., the tendency to agree with test items regardless of test item content). The Acquiescence scale was scored as the mean difference between the test item pairs. Hand also developed a social desirability scale which was comprised of item pairs from the Guilford-Zimmerman Temperament Survey. Item pairs were chosen such that one item was socially desirable and the other item was socially undesirable, and both items were scored alike (e.g., both items were keyed in the positive direction). Furthermore, both items measured the same trait. The resultant scale contained 40 items and was scored as the mean difference between the item pairs. The Social Desirability and Acquiescence scales displayed a negligible relationship ($r = -.05$). Furthermore, the Social Desirability

scale demonstrated meaningful relationships ($r = .54$) with the Edwards Social Desirability Scale (Edwards, 1957), and the Acquiescence scale demonstrated a moderate relationship ($r = .36$) with Bass' Social Acquiescence scale (Bass, 1956).

Researchers investigating consistency-in-responding have developed two approaches to detecting consistency-in-responding, namely using identical items (Buechley & Ball, 1952) and using items that represent opposite valences within the same test (Greene, 1978). Both approaches use item pairs (i.e., two items that are either identical or similar in content) and calculate response inconsistency as the difference between responses on the pairs of items (e.g., absolute difference). Researchers have used this approach with several scales, including scales developed for normal populations (e.g., NEO-PI-R; Kurtz & Parrish, 2001) and clinical populations including the MMPI (Wetter, Baer, Berry, Smith & Larsen, 1992) and the Personality Assessment Inventory (PAI; Morey, 1991). The most widely studied consistency-in-responding scale is the Variable Response Inconsistency (VRIN) scale from the MMPI. Some researchers argue that VRIN is only sensitive to random responding (e.g., Berry, Wetter, Baer, Larsen, Clark, & Monroe, 1992; Tellegen, 1982, 1988). Thus, practitioners may examine VRIN scale scores along with other indicators of random responding (e.g., True Response Inconsistency [TRIN] scale scores) and discard substantive scale scores for test-takers who exceed some predetermined scale score on the consistency-in-responding scale. That is, if these indicators are interpreted as evidence of random responding, substantive interpretation of the scales may be inappropriate. The VRIN scale appears to be sensitive to wide-scale random responding (i.e., responding

randomly to all or the majority of test items) but its effectiveness in detecting partial random responding (i.e., responding randomly to a small subset of test items) is not well understood (Clark, Gironda, & Young, 2003).

It is important to note that researchers vary in their interpretation of consistency–in–responding scores. Some researchers view inconsistent responding as a response style as opposed to a response set. Response styles are tendencies to respond to test items regardless of their content, whereas response sets are content–specific. Response distortion is typically conceptualized as a response set. Bond (1986) questions the assumption that inconsistent responses to similar items are only due to random or carelessness in responding. Several plausible alternative causes of inconsistency include indecision, response distortion, and acquiescence. Furthermore, inconsistency can be caused by item characteristics (e.g., item ambiguity, item subtlety) and interactions between the substantive causes (e.g., indecision) and item characteristics (Bond, 1987). Several researchers (e.g., Edwards & Walsh, 1963; Goldberg, 1963; Payne, 1974) found that items with high social desirability values elicit more consistent responding compared to items with moderate social desirability values. Similarly, Bond (1987) argues that inconsistency–in–responding is best attributed to response indecision, which may be caused by item ambiguity, perceived personal applicability of test item content, and conflict regarding self–disclosure in admitting to undesirable traits. Furthermore, inconsistency–in–responding scale scores tend to be positively related to MMPI F scale scores, where MMPI F scores indicate response distortion (Bond, 1986; see Buechley & Ball [1952] for an alternative perspective).

In summary, it would appear that the interpretation of inconsistent responses generally differs depending on whether the scale was developed for normal or abnormal populations. In clinical settings, it is posited that test-takers may engage in random responding for a variety of reasons (e.g., carelessness, indecision). One potential cause of random responding that seems unique to clinical settings is using random responding to obscure one's true standing on the trait of interest. That is, some test-takers may be reluctant to respond honestly to scales that measure sensitive personal attributes, and rather than responding in an overly desirable or undesirable manner, they simply respond randomly. Conversely, inconsistency-in-responding can be interpreted as a response set. For example, test-takers who are attempting to distort their responses may have difficulty responding consistently to item pairs when one of the items is socially desirable, and one is socially undesirable.

Another approach to response distortion detection focuses on the social desirability of *response options* as opposed to the social desirability of *test items*. Specifically, Kuncel and Tellegen (2009) posit that a focus on the social desirability of item responses may be more appropriate than an examination of the social desirability of items, as initial evidence suggests that moderate response options are sometimes perceived as more desirable than extreme response options. For example, being moderately assertive may be perceived as more desirable than being extremely assertive or extremely passive. As such, Kuncel and Borneman (2007) developed an idiosyncratic approach which uses personality test items which display bi- or tri-modal response distributions in an instructed-faking condition but display relatively normal response

distributions in an instructed–honesty condition. The theoretical and conceptual basis for using items that display these properties was that these items were ambiguous as to the level of the trait that was most desirable. This approach correctly identified 70.0% of test–takers in the instructed–faking condition and incorrectly identified 17.5% of test–takers in the instructed–honest condition as faking. Furthermore, this new approach displayed minimal correlations with an improbable item lie scale, and smaller correlations between the Five Factor Model (FFM) personality dimensions and general mental ability (GMA) compared to an improbable item lie scale. As such, this new approach appears to hold some promise.

In addition to lie scales, researchers have investigated new and innovative methods for detecting response distortion including the use of response latencies and eye–tracking. Response latency researchers argue that the time between the presentation of a personality test item and the test–taker’s response may be an indicator of response distortion. However, there is a continued debate as to whether response distortion results in longer or shorter response latencies. Some researchers posit that engaging in response distortion is more cognitively complex than responding in an honest manner (e.g., Zuckerman, DePaulo, & Rosenthal, 1981), which in turn results in longer response latencies. Conversely, other researchers argue that test–takers engaging in response distortion do not process items in the same manner as honest test–takers (Hsu, Santelli, & Hsu, 1989; Holtgraves (2004). That is, engaging in response distortion may allow test–takers to skip stages in the response process, thus resulting in shorter response

latencies. Unfortunately, it is currently quite unclear whether response distortion engenders longer or shorter response latencies compared to honest responding.

Another innovative method for detecting response distortion involves the use of eye-tracking data. Eye-tracking research investigating response distortion involves documenting where test-takers fix their gaze (i.e., fixate) when reading personality test items. When reading test items, test-takers alternate between rapid eye movements, called saccades, and moments when the eyes are relatively still (i.e., fixations).

Fixations are indicators of interest. Van Hooft and Born (2012) argued that the number of fixations would be an indicator of cognitive processing. As such, relatively more fixations during an instructed-faking compared to instructed-honesty would indicate that response distortion was more cognitively demanding than responding honestly.

Conversely, more fixations during an instructed-honesty condition compared to a instructed-faking condition would indicate responding honestly was more cognitively demanding than engaging in response distortion. Van Hooft and Born found that response distortion is less cognitively demanding with test-takers having less fixations during the instructed-faking condition compared to the instructed-honesty condition. Additionally, Van Hooft and Born found that in the instructed-faking condition, test-takers fixated more on the 2 extreme responses, and that these fixations occurred more temporally proximal to when the test-takers finished reading the item stem compared to the instructed-honesty condition.

In summary, the goal of lie scales is to identify test-takers who may have distorted their responses. Test-takers who score above some predetermined threshold on

the lie scale are flagged, and their personality test scores may be corrected or invalidated. These scales are useful in identifying a subset of individuals who distort their responses, although there is evidence that some test-takers can engage in response distortion without detection. Furthermore, lie scales can take a number of approaches to detecting distortion (e.g., improbable or bogus items), and their efficacy appears to vary as a function of the specific approach used.

Contrasting Groups. As indicated in several of the studies reviewed in the preceding sections, researchers often use an instructed-faking approach to assess the susceptibility of noncognitive measures to response distortion (e.g., Converse, Oswald, Imus, Hedricks, Roy & Butera, 2008; Kuncel & Borneman, 2007; McFarland, Ryan, & Ellis, 2002; McFarland & Ryan, 2006; Viswesvaran & Ones, 1999). This approach allows researchers to assess the susceptibility of noncognitive measures to response distortion, make comparative evaluations between measures, and evaluate the efficacy of newly developed techniques for deterring or detecting response distortion. For example, McFarland and Ryan (2000) evaluated the magnitude of response distortion across three noncognitive measures (i.e., personality, biodata, and integrity). Their results suggest that, even under instructed-faking, there were considerable individual differences in response distortion and that these individual differences were fairly consistent across measures. Furthermore, their results suggest that the FFM dimensions are differentially susceptible to response distortion. Specifically, conscientiousness and emotional stability displayed difference scores approximately twice as large as that of extroversion and agreeableness, and ten times that of openness (cf. Viswesvaran & Ones, 1999).

Although there seems to be consensus that test-takers can distort their responses (e.g., Viswesvaran & Ones, 1999), the effects observed in the laboratory are markedly larger than that generally found in field settings (see Mitchell, 2012). Instructed-faking typically results in a one standard deviation increase in observed scores compared to instructed-honesty conditions (Viswesvaran & Ones, 1999), whereas applicant versus incumbent comparisons typically display one-half of a standard deviation difference (with applicants having higher mean scores; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006).

It is worth noting that instructed-faking paradigms can take many forms including so-called “fake-good”, “fake-bad”, and “fake-job” instructions or approaches. Fake-good instructions ask test-takers to complete a personality measure in such a way as to give the best possible impression. Fake-bad instructions ask test-takers to complete a personality measure in such a way as to give an unfavorable impression. Both “fake-good” and “fake-bad” approaches can vary in terms of the specific instructions used. For example, Paulhus, Bruce, and Trapnell (1995) examined the effects of six different faking instructions, which varied in terms of both their overtness (e.g., fake best, fake modest) and valence (fake good, fake bad). The results demonstrated that both overtness and valence resulted in different patterns of mean scores. However, none of the instructions specified a target job. Conversely, fake-job instructions ask test-takers to complete a personality measure in such a way as to increase their chances of being hired if they were job applicants for a specific job (e.g., psychiatric nurse; Mahar, Cologn, & Duck, 1995).

In the context of selection testing, the extent to which fake-good and fake-bad instructions are appropriate is unclear (Mahar et al., 1995). Specifically, fake-good and fake-bad instruction sets are used to investigate response distortion in both personnel and clinical assessment. In the context of clinical assessments, test-takers may distort their responses to create a favorable impression to avoid aversive interventions (e.g., concealing aggression to avoid an anger management intervention) or distort their responses to create an unfavorable impression in an attempt to receive pleasant interventions (e.g., malingering to elicit attention and sympathy) or avoid unpleasant outcomes (e.g., faking mental illness to invoke an insanity defense in a legal trial). However, in selection contexts, it is likely that actual applicants would distort their responses to increase their chances of obtaining employment rather than to give a generally good impression. Furthermore, it is difficult to envisage circumstances in which a test-taker would fake-bad on a selection test, since selection tests typically are not compulsory.

Although the fake-job approaches to investigating response distortion appear more germane to selection contexts, relatively few studies have used this approach. The majority of these studies (e.g., Furnham, 1990; Kaufman, Hakmiller, & Porter, 1959; Kroger & Turnbull, 1975; Velicer & Weiner, 1975) have found that respondents are able to distort their responses to stereotypic profiles of salient jobs (e.g., Air Force officer), though not ambiguous jobs (e.g., creative artist). However, the author is unaware of any studies that have investigated the extent to which respondents can distort their responses to match an ideal profile (i.e., a profile associated with high performance).

Although the extant literature demonstrates that most, if not all, noncognitive measures are susceptible to response distortion to some degree, and that applicants typically score higher than incumbents on self-report measures (e.g., Birkeland et al., 2006; Bott, O'Connell, Ramakrishnan, & Doverspike, 2007), there is considerable disagreement in the literature regarding the extent of response distortion in applied settings and the effects of said distortion.

Response Distortion Debate

Some researchers suggest that response distortion on self-report noncognitive tests is unavoidable and so these tests should not be used in high-stakes settings. Morgeson et al. (2007a, 2007b) argue that response distortion cannot be avoided, and response distortion, coupled with relatively low observed criterion-related validity, threatens the operational use of personality tests in high-stakes settings. Furthermore, they argue that if personality tests are going to be used, then off-the-shelf self-reported personality tests should be avoided in lieu of either personality measures that are not self-report or customized personality measures. The use of customized personality measures would ostensibly increase criterion-related validity, increase face validity, and deter response distortion. Indeed, customized personality measures may display lower levels of response distortion compared to published measures, as customized measures may have lower levels of item exposure. However, the potential benefits of customized personality measures (e.g., increased criterion-related and face validity, less susceptibility to response distortion) have not received empirical attention.

In response to Morgeson et al. (2007a), Tett and Christiansen (2007) argue that response distortion occurs in applicant settings and that it attenuates the observed criterion-related validity coefficients. In contrast to Morgeson et al., they conclude that said reduction does not diminish the criterion-related validity of personality test scores to the point where they are useless for personnel decision-making. Tett and Christiansen suggest future research could both reduce the extent and effects of response distortion and increase the observed relationship between personality test scores and performance outcomes. Specifically, they recommend taking a more process-oriented and theory-driven approach to investigating the relationship between personality and job performance as well as the extent and effects of response distortion. Similarly, Hough and Oswald (2008) recommend systematically investigating the role of test-taking mode, testing context, instruction, individual differences, external motivating factors, and test-taking outcomes in response distortion (see also McFarland & Ryan, 2006; Tourangeau & Yan, 2007).

Other researchers argue that response distortion may occur; however its impact is minimal (e.g., Barrick & Mount, 1996; Cunningham, Wong, & Barbee, 1994; Ellingson, Smith, & Sackett, 2001; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Ones, Dilchert, Viswesvaran, & Judge, 2007). Specifically, Hough et al. (1990) argue that response distortion has a negligible impact on the criterion-related validity evidence of personality measures. They measured 10 personality constructs and two personality-relevant job performance outcomes resulting in 20 criterion coefficients. Only 4 out of 20 coefficients displayed differential relationships between the personality predictor and

performance measure as a function of response distortion. Ellingson, Sackett, and Connelly (2007) used a repeated measures design where test-takers initially completed a personality measure for developmental or selection purposes, and then retested for selection or developmental purposes, creating four conditions (i.e., developmental/selection, selection/developmental, developmental/developmental, and selection/selection). This design allowed for the score changes to be decomposed into components attributable to retesting, feedback, and response distortion effects. When the effects of retesting and feedback are taken into consideration, the average response distortion effect was quite small ($d = .08$).

Finally, other researchers argue that, although response distortion may not affect criterion-related validity, response distortion may impact hiring decisions (Christainsen, Goffin, Johnston, & Rothstien, 1994; Rosse, Stecher, Miller, & Levin, 1998; see also Bott et al., 2007). Hiring decisions may be affected by response distortion depending on the distributional placement of test-takers who engage in response distortion. Specifically, response distortion will only influence hiring decisions when test-takers differentially engage in response distortion (i.e., some test-takers distort their responses more than others), and when response distortion occurs in the upper end of the distribution. Using a repeated measures design, Arthur et al. (2010) found that test-takers who distorted their responses were generally evenly distributed across the score range, with a slight trend of more response distortion occurring in the upper score range. Christainsen et al. (1994) compared the hiring recommendation made on the basis of personality test scores that were uncorrected and corrected for response distortion (as

measured by two lie scales) at varying selection ratios. In general, the effect of response distortion on hiring recommendations varied as a function of the selection ratio, such that the discrepancy between the corrected and uncorrected scores increased as the selection ratio decreased. At a selection ratio of .15, 16% of the hiring recommendations were discrepant. That is, if the uncorrected personality test scores were used, 16% of the test-takers would have obtained their employment via response distortion by displacing honest test-takers (cf. Schmitt & Oswald, 2006).

The focus on individual decisions made on the basis of personality test scores (or any other selection device) has been critiqued because the utility of personnel selection tools are predicated on incremental gains over a large number of decisions (Schmitt & Oswald, 2006). Specifically, the standard errors associated with selection tools do not provide accurate prediction for individual decisions. As such, the traditional focus has been on group performance. However, there are important implications and concerns with using tests which allow test-takers to engage in malfeasant behavior and to gain valued rewards by doing so. Specifically, the use of suspect selection tools may negatively impact recruitment and offer acceptance when test-taker reactions are poor (Hausknecht, Day, & Thomas, 2004). Furthermore, there are ethical implications associated with “liars” displacing honest respondents via response distortion (see Morgeson et al., 2007a, 2007b where the authors take contrasting perspectives).

Item Exposure and Coaching Threats

Although there is considerable disagreement about the extent to which response distortion threatens the operational use of personality tests, the widespread use of

personality tests has led to additional threats. Specifically, large-scale testing and the use of unproctored internet-based testing has led to item exposure and coaching concerns. For example, in 2008 the Wall Street Journal reported that a website had posted operational test items of the Graduate Management Admission Test. As a result, 84 test results were invalidated (Hechniger, 2008). Similarly, the answer key to a personality test developed by a large testing firm was posted on the internet (O'Connell, 2009). Specifically, this testing firm provided a personality test for a number of national retail chains in an unproctored internet-based manner. Because the same test is used for a number of large organizations, it is possible that test-takers can take the test repeatedly, discuss test content and response strategies with others, and find test items and "cheat sheets" online. A number of people interviewed for the report admitted to engaging in said behaviors. Furthermore, there is a weblog (or blog) dedicated to discussions regarding the specified personality test and hosts an answer sheet. Additionally, Burke (2009) reported that 18 internet sites offered test content from a single testing firm. Finally, books on how to "beat" personality tests are widely available (Whyte, 1975).

In addition to test security and item exposure concerns, the extent of response distortion may also be influenced by test coaching. Although test coaching has traditionally focused on cognitive ability tests (e.g., Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007), coaching has also been investigated in the context of integrity tests (Alliger & Dwight, 2000; Alliger et al., 1996; Hurtz & Alliger, 2002), interviews (Maurer, Solamon, & Troxtel, 1998; Maurer, Solamon, Andrews, & Troxtel, 2001),

situational judgment tests (Cullen et al., 2006), assessment centers (Sackett, Burris, & Ryan, 1989), and self-reported personality tests (e.g., Zickar & Robie, 1999). The majority of research on response distortion in general, and coached response distortion in particular, has been conducted using the MMPI (Storm & Graham, 2000). The MMPI contains lie scales to detect faking good (i.e., defensiveness) and faking bad (i.e., malingering). The preponderance of coaching studies have focused on faking bad (e.g., Lamb, Berry, Wetter, & Baer, 1994; Rogers, Bagby, & Chakraborty, 1993; Storm & Graham, 2000) and have found that coaching increases the likelihood that test-takers can malingering without being detected. Zickar and Robie (1999) investigated the efficacy of an Item Response Theory-based detection technique to identify test-takers who completed the Assessment of Background and Life Experiences (ABLE; Hough et al., 1990) under one of three instruction sets—instructed-honesty, instructed-faking, and instructed-faking with coaching. Although coaching was not the primary focus of their study, their results suggested that coaching resulted in higher scores compared to the instructed-honesty and instructed-faking conditions.

Miller and Barrett (2008) argue that the efficacy of personality and other noncognitive measures is particularly threatened, due to distortion and coaching concerns, when they are used to select municipal safety forces (e.g., police officers). Specifically, the presence of unions and a particularly litigious environment result in widespread coaching concerns. Miller and Barrett document several cases where for-profit firms provided test-takers with “hypothetical” test items which were similar or identical to operational test items during training sessions that only a subset of test-

takers attended. These for-profit firms also offer training regarding methods of deterring and detecting response distortion, and how to “beat” these systems. As such, test-takers who are motivated to do so, and are able to pay the training fees, greatly increase their chances of effectively distorting their responses without detection and thus increase their chances of being hired.

Although there is relatively little research regarding coaching on personality tests, there have been at least five articles (i.e., Ellingson et al., 2007; Hogan, Barrett, & Hogan, 2007; Kelley, Jacobs, & Farr, 1994; Landers et al., 2010; Young, 2003) that examined the extent of response distortion using a repeated measures design where test-takers were ostensibly motivated to engage in response distortion during initial testing and upon retest. Two of these studies (Ellingson et al., 2007; Hogan et al., 2007) found small differences whereas three studies found comparatively larger differences (i.e., Young, 2003; Landers et al., 2010; Kelley et al., 1994). For example, Kelly et al. (1994) investigated MMPI test scores from nuclear power workers over repeated administrations. Five out of 11 scales displayed statistically and practically significant increases over administration. The authors argue that said increases may be attributed to response distortion, but the data did not allow for an examination of this supposition. Furthermore, an examination of individual-level score changes suggested that, even for scales that displayed small mean differences, a substantial proportion of the sample displayed meaningful score changes across administrations (although in different directions). Finally, scale scores in a critical range (i.e., scale scores that might screen-out test-takers) decreased across administration. Specifically, the majority (59.9%) of

test-takers were flagged as potentially problematic, whereas 36.6% were flagged upon a second administration. Thus, the researchers concluded that the efficacy of this test was threatened by repeated administrations.

Effects of Response Distortion

Researchers have investigated the effects of response distortion on the psychometric properties of personality test scores, including observed means, and construct- and criterion-related validity. A large body of literature indicates that applicants display higher scores compared to incumbents. Birkeland et al. (2006) investigated job applicant response distortion by comparing applicants' and incumbents' scores on the FFM personality dimensions. Their results suggest that the sample-weighted mean difference between applicants and incumbents varied as a function of the personality dimension and whether the personality test was designed to measure the FFM (i.e., direct measure) or was retrofitted into the FFM framework (i.e., indirect measure). Specifically, for all personality dimensions, direct measures resulted in larger differences compared to indirect measures. Also, conscientiousness ($d = -0.79$) and emotional stability ($d = -0.72$) displayed the largest mean differences followed by agreeableness ($d = -0.51$), openness ($d = -0.28$), and extraversion ($d = -0.18$) for direct measures. Thus, the proposition that applicant means tend to be higher than incumbent means is approaching the status of a received doctrine.

Construct-Related Validity. The factor structure of FFM measures of personality (e.g., NEO-PI) has received considerable research attention (McCrae & Costa, 1987, 1989, 1990) using volunteer samples. These investigations have generally supported the

FFM model. However, a critical question is whether the factor structure is invariant across testing contexts (e.g., research setting versus selection setting). Schmit and Ryan (1993) compared the factor structure of personality tests using a student and applicant sample. Their results suggested that the FFM factor structure adequately described the student sample data, but not the applicant data. Specifically, the applicant data resulted in an additional factor which they labeled an "ideal–employee" factor. Subsequent studies have found similar factor structures for FFM measures in applied settings (e.g., Cellar, Miller, Doverspike, & Klawnsky, 1996).

Relatedly, the intercorrelations between personality factors increase under instructed–faking, as do the intercorrelations between lie scales and substantive personality dimensions (Gorman, 1968; Rump & Court, 1971; Stanush, 1996). Thus, response distortion affects not only the observed score means, but also the factor structure, and thus the construct–related validity of personality test scores.

Criterion–Related Validity. A primary concern of researchers and practitioners is the extent to which response distortion influences the criterion–related validity of personality test scores in predicting job performance. Several studies have examined the extent to which response distortion influences criterion–related validity using lie scales (e.g., Hough et al., 1990), and contrasting groups (i.e., applicants versus incumbents, instructed–faking versus instructed–honesty). Some of these studies have concluded that the criterion–related validities of test scores are not affected by response distortion (e.g., Ellingson et al., 2001; Hough et al., 1990; Ones et al., 1996; Smith & Ellingson, 2002),

whereas other studies have found that they are (Dunnette, McCartney, Carlson, & Kirchner, 1962; White et al., 2001).

There are several possible reasons for these discrepant findings. First, response distortion appears to attenuate the criterion-related validity in instructed-faking designs (Holden & Jackson, 1981; Schmit & Ryan, 1992) to a greater extent than in applied settings (Ellingson, Sackett, & Hough, 1999). Furthermore, when lie scales are used to detect response distortion, the extent to which criterion-related validities are attenuated appear to vary as a function of the type of lie scale used. Specifically, Hough et al. (1990) used an improbable item approach and found little to no attenuation, whereas Pannone (1984) used a bogus item approach and found attenuation. Finally, studies vary in terms of the presence or absence of techniques to deter response distortion (e.g., forced-choice response formats, warnings). Given the general lack of theory regarding the antecedents and consequences of response distortion, it is difficult to generate summary statements regarding the effect of response distortion on the criterion-related validity of personality tests. To this end, McFarland and Ryan (2000) proposed an integrated model of response distortion. However, few studies have systematically investigated response distortion using this framework (e.g., McFarland & Ryan, 2006).

Several meta-analyses have examined the effect of response distortion on the criterion-related validity of personality test scores. Li and Bagger (2006) investigated the extent to which response distortion (as measured by the BIDR version 6; Paulhus, 1991) attenuated the criterion-related validity of personality test scores. Their results

suggest that neither impression management nor self-deceptive enhancement meaningfully suppressed or moderated the criterion-related validities. For example, partialing out impression management from agreeableness scores resulted in a small decrease in validity (unadjusted $r = 0.13$, adjusted $r = .09$). Another meta-analysis by Ones et al. (1996) concluded that response distortion (as measured by lie scales) does not function as a suppressor or moderator variable. However, the veracity of this conclusion depends on the assumption that lie scales accurately capture response distortion.

Several other meta-analyses have examined the criterion-related validity of personality test scores in both concurrent (incumbent) and predictive (applicant) designs (e.g., Hough, 1998; Tett, Jackson, & Rothstein, 1991). The results of Hough's (1998) meta-analysis suggested personality test scores display meaningful relationships with outcomes of interest in both concurrent and predictive validity studies. The effect size for concurrent studies was .07 points higher than the effect size for predictive studies. Tett et al.'s meta-analysis found quite similar results. There are two competing explanations for this observed difference. First, the time lag associated with predictive studies could explain the smaller effect size. Second, a response distortion hypothesis would postulate that response distortion introduced systematic error variance into the personality test scores which attenuated the relationship. Unfortunately, the data presented could not rule out either possibility.

Techniques to Detect and Deter Response Distortion

Although the effects of response distortion remain unclear, the preponderance of the extant literature suggests that response distortion occurs when personality tests are

used in high–stakes testing (Arthur et al., 2010; Birkeland et al., 2006; Griffith et al., 2007; Landers et al., 2011; Levin & Zickar, 2002; cf. Hogan et al., 2007; Hough et al., 1990). Thus, several techniques for mitigating response distortion have been proposed. These techniques can be categorized into methods for deterring, and methods for detecting response distortion. Techniques for deterring response distortion include empirical keying, forced–choice response, and warnings. Methods of detection include the use of lie scales, verification, and response elaboration (see Hough, 1998 for a review). It is important to note that some of these techniques may be used to both detect and deter response distortion, especially if they are used in combination. For example, warnings of response elaboration may serve to deter response distortion whereas response elaboration is a detection technique. Although some of these techniques hold some promise, a majority of these techniques are either ineffective or have limitations that reduce their efficacy in applied settings.

Warnings. Warnings typically take the form of instructions or interactive prompts that inform test–takers to respond to personality items honestly. They inform test–takers that response distortion can be detected, thus increasing the fear of detection (McFarland & Ryan, 2000). Landers et al. (2011) found that interactive warnings reduced the observed extent of response distortion in the form of blatant extreme responding (i.e., responding using only the most extreme response options). Warnings appear to be particularly effective when consequences are attached to score inflation (Dwight & Donovan, 2003; McFarland, 2003). However, warnings may cause otherwise honest respondents to “fake bad” in response to the warnings (Oswald & Hough, 2010).

Furthermore, warnings of verification led to higher correlations between GMA and conscientiousness compared to when there was no warning (Vasilopoulos, Cucina, & McElreath, 2005). Finally, the efficacy of warnings depends on test-takers being willing to respond in an honest manner.

Forced-Choice Response Formats. Forced-choice response formats have also received attention in the extant literature (e.g., Chernyshenko, Stark, Prewett, Gray, Stilson, & Tuttle, 2009; Dunnette et al., 1962; Heggstad, Morrison, Reeve, & McCloy, 2006). Forced-choice personality measures instruct test-takers to endorse one of two response options which measure different personality traits. During test construction items are paired in terms of their desirability (i.e., items of equivalent desirability are paired together) where item desirability is typically operationalized as the observed endorsement rates when the items are presented separately. For example, test-takers might have to choose between endorsing a desirable conscientiousness statement or a desirable emotional stability statement. The results of a forced-choice personality measure indicate the relative strength of the personality constructs measured.

Forced-choice response formats appear to be partially resistant to response distortion (Christiansen, Burns, & Montgomery, 2005; Vasilopoulos, Cucina, Dynomina, & Morewitz, Reilly, 2006; cf. Heggstad et al., 2006). However, their development is often resource-intensive, they do not entirely mitigate response distortion, they result in ipsative scores, and their construct-related validity is questionable. Specifically, forced-choice response formats typically result in total test scores that are equal across test-takers (i.e., ipsative scores). Ipsative scores force dependence among responses which

impose psychometric problems including difficulty in assessing factor structure (Dunlap, & Cornwell, 1994), reliability (Tenopyr, 1988), and validity (Meade, 2004).

Furthermore, the use of forced-choice response formats engender different response processes that lead to an inflation of the relationship between personality test scores and GMA (Christiansen et al., 2005; Vasilopoulos et al., 2006).

In summary, researchers and practitioners continue to be concerned with response distortion on self-report noncognitive tests. Although there is agreement that response distortion can occur, there is an ongoing debate regarding the extent, magnitude, and effects of response distortion in applied settings. Using lie scales and contrasting groups are two methods for investigating response distortion. Unfortunately, results regarding response distortion often vary as a function of the types of methods used. The lack of agreement regarding the effects of response distortion provides impetus for developing new methods for deterring and detecting response distortion. Although a subset of these methods show some potential (e.g., warnings, forced-choice response formats), there are concerns and disadvantages associated with their use in selection contexts.

One method of mitigating response distortion that has not received much research attention is the use of profile matching. Profile matching is a personnel decision-making strategy that entails assessing the fit between an ideal or standard profile and test-taker scores. Rather than focusing on the magnitude of test scores, as is the case with linear models, profile matching focuses on the configuration of test scores. That is, the most desirable score is not necessarily at an extreme end of the scale, and the

desired level of a specific trait depends on the levels of the other traits contained in the profile. Thus, conceptually, preknowledge of test items and coaching threats are not a concern as long as the standard profile is not known to the test-takers.

CHAPTER III

PROFILE MATCHING

In personnel decision-making there are four widely recognized strategies for making decisions based on multiple predictors, namely multiple regression, multiple cutoff, multiple hurdle, and profile matching. Three of these strategies—multiple regression, multiple cutoff, and multiple hurdle—are based on linear models which, given a positive relationship between predictors and criteria, assume that extreme high predictor scores are always associated with higher criterion scores and thus, higher predictor scores are always desirable and preferable. However, some personality researchers (e.g., Chaplin, 2007) argue that linear prediction models are oversimplifications, as human behavior, and specifically job performance, is thought to be quite complex (Campbell, Gasser, & Oswald, 1996). Furthermore, although the use of profile matching is somewhat limited in I/O psychology, clinical and counseling psychology have long recognized the importance of using profile matching as a decision-making strategy in the diagnosis and treatment of mental illness (Patrick, 1984; Wirt et al., 1984).

The reliance on linear models in personnel decision-making involving noncognitive predictors may exacerbate problems associated with response distortion, especially in the context of top-down selection (Arthur et al., 2001). The propensity to use selection strategies that are based on linear models in lieu of profile matching may be due, in part, to the conceptual and analytical challenges engendered in nonlinear research. In detailing issues associated with nonlinear research, Chaplin (1997)

discusses several challenges. The first issue was well articulated by Cronbach (1957), “Once we attend to interactions, we enter a hall of mirrors that extends to infinity. However far we carry our analysis—to third order or fifth order, or any other order—untested interactions of a still higher order can be envisioned” (p. 119). That is, potential moderators of any particular predictor–criterion pair may be quite numerous, and the effect size of any particular moderator is likely to be quite small.

Furthermore, the statistical power for detecting an interaction effect is, by definition, lower than that of the main effects and the overall statistical power of an analysis is generally inversely related to the number of predictors in a model. Thus, from a statistical perspective, it may be quite difficult to detect moderator effects unless the effect size or sample size is quite large. From a psychometric perspective, including interaction effects involves adding terms (i.e., cross products) that have relatively small reliabilities, as the maximum reliability of a cross product is the product of each component’s reliability assuming the components are not correlated. These decrements in reliability are associated with decrements in validity, as reliability forms the upper bound for validity.

The final challenge in conducting nonlinear research discussed by Chaplin (1997) is building a conceptual framework that specifies the moderator effects of interest and drives measurement and research design decisions. That is, efforts to discover moderator effects post hoc are likely to be unsuccessful. Instead, researchers should use theory in selecting the proposed moderator variable, clearly delineate the mechanism by

which the moderator influences the relationship between the independent and dependent variable, and specify the effect of the moderator on said relationship (Chaplin, 2007).

In addition to the challenges discussed by Chaplin (1997) which focus on the difficulties faced by a researcher interested in nonlinear effects, Murphy (1996) argues that I/O researchers are indoctrinated into thinking about simple linear relationships almost exclusively. Specifically, early I/O research focused on demonstrating the role of GMA in performance (Hunter, 1986; Hunter & Hunter, 1984; Ree & Earles, 1992). Since the GMA–performance relationship is linear (Coward & Sackett, 1990), the relationship can be described by a correlation coefficient. As such, Murphy argues that I/O researchers may be “victims of our own success” (p. 6) insofar as the tradition of the GMA–performance research leads to oversimplified research questions and statistical approaches (cf. Cohen, 1990). That is, the dominance of linear models and their associated statistics (e.g., correlation coefficients) may restrict the proliferation of theories that include more complex, nonlinear models.

Although personnel selection is characterized by the use of linear models, several researchers question the use of linear models in personnel decision–making especially when personality variables are used as predictors (e.g., Arthur et al., 2001; Murphy, 1996; Murphy & Dzieweczynski, 2005; Ones et al., 2007). Arthur et al. argue that the assumption of linearity may be appropriate for some predictors (e.g., knowledge, skill, ability), but probably not personality since unlike cognitive ability, which displays a strong positive manifold (e.g., Ackerman & Humphreys, 1990), personality variables display negligible to moderate intercorrelations (Ones et al., 1996). Tests of knowledge,

skills, and ability typically yield scores that show a consistent pattern of positive correlations among these tests and between the tests and criteria (Allinger, 1988; Humphreys, 1979; Ree & Carretta, 2002). Murphy, Dzieweczynski, and Yang (2009) argue and present evidence that supports the claim that positive manifold is the norm rather than the exception for a number of predictor constructs including GMA, psychomotor ability, and spatial ability, and a number of predictor methods including interviews, biodata, work samples, and situational judgment tests (see also Murphy, 2009). Personality measures appear to be somewhat unique in the fact that they do not display positive manifold. If the relationships between specified personality variables and performance are nonlinear, then decision-making strategies that are based on a linear model (e.g., multiple regression, multiple cutoff, and multiple hurdle) maybe inappropriate.

Profile matching is a decision-making strategy that entails assessing the fit between candidate test scores across multiple predictors and some ideal or standard profile (see Figure 1 for an illustrative example of score profiles). As such, profile matching does not assume linearity between predictor scores and criteria. So, this approach does not assume that extreme predictor scores are always associated with higher criterion scores. Rather, a desirable score is one that matches or minimally deviates from the ideal score. As the observed score deviates from the ideal score, the associated forecasted criterion score decreases. As such, profile matching differs from the typical bivariate treatment of personality variables in that it can capture nonlinear effects and takes a multivariate, whole-person approach.

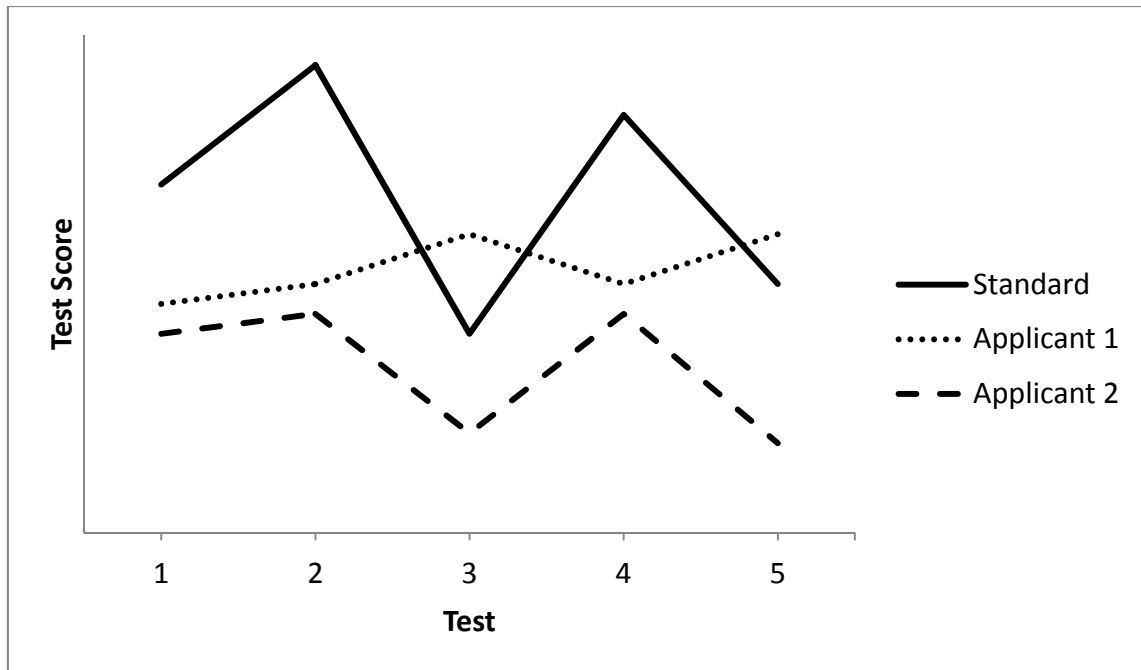


Figure 1. An illustrative example of 3 score profiles.

Nonlinear Effects

Nonlinear effects can take multiple forms, however the preponderance of research has investigated possible quadratic, and to a lesser extent cubic, models. Two forms of quadratic models are particularly germane to a discussion of personality and performance—asymptotic and inverted-U models. An asymptotic model suggests that beyond some threshold, increases in predictor scores are no longer associated with increases in criterion scores (see Figure 2). From this perspective, although there are no decrements in performance associated with selecting candidates at the extreme end of the distribution, any candidate past a specified threshold would be expected to perform as well as any other candidate whose score is more extreme than the threshold. Indeed, several researchers have posited that the relationship between specified personality

variables and outcomes of interest may be best described by an asymptotic model (e.g., Barrick & Mount, 1991; Vasilopoulos, Cucina, & Hunter, 2007). For example, Barrick and Mount evoked a threshold (asymptotic) argument when they proposed that there may be a critical range for emotional stability, such that the emotional stability–performance relationship is minimized for individuals who possess an adequate level of emotional stability.



Figure 2. An illustrative example of an asymptotic relationship.

An inverted–U model suggests that deviations from an optimal score are associated with performance decrements, with said deviations occurring at the extreme ends of the distribution (see Figure 3). Selecting candidates at the extreme positive end of the distribution is associated with lower levels of performance. The Yerkes–Dodson law (Yerkes & Dodson, 1908) is a typical example of an inverted–U relationship. Specifically, the Yerkes–Dodson law states that moderate levels of arousal are

associated with relatively higher levels of performance compared to both extremely low and extremely high levels of arousal. Benson and Campbell (2007) posited that derailing traits would be nonlinearly related to leadership performance. They found evidence for an inverted-U relationship between these personality variables and leadership performance in two samples using different sets of self-reported personality measures. Several researchers have argued that conscientiousness (e.g., Le et al., 2010; Murphy, 1996) and emotional stability (Le et al., 2010) may be best described using an inverted-U model. Furthermore, Day and Silverman (1989) reported that impulse expression was related to timeliness of work (a facet of job performance) such that incumbents who were high and low on impulse expression were rated lower in terms of timeliness of work than those who were moderate in impulse expression.

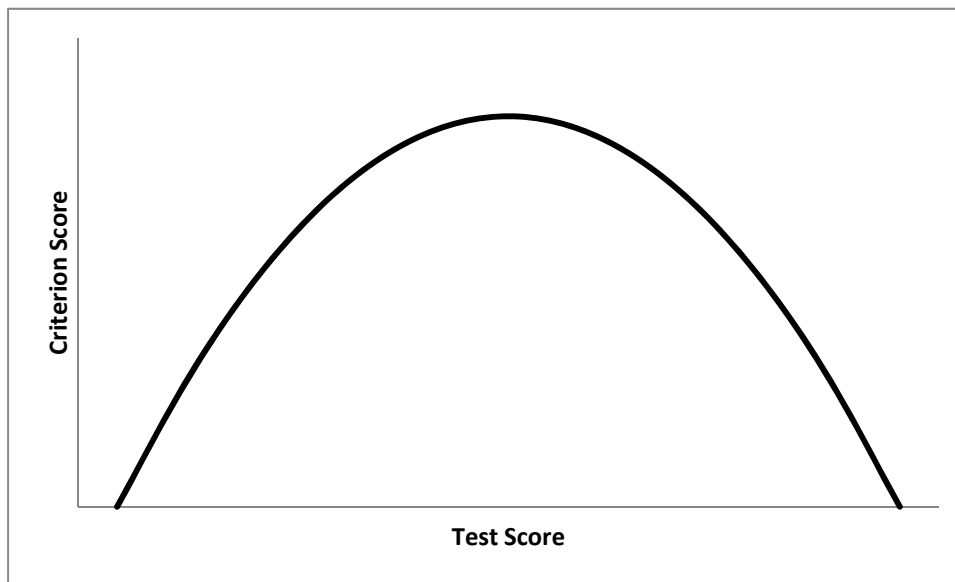


Figure 3. An illustrative example of an inverted-U relationship.

A cubic model suggests that low predictor scores are associated with low criterion scores and high predictor scores are associated with high predictor scores, but scores in the middle range of the distribution result in the same predicted criterion scores (see Figure 4). The strength of the situation (Mischel, 1977) could suppress expression of personality such that only individuals who are extremely high or extremely low on a trait will express that trait (Robie & Ryan, 1999). For example, in a department that has a strong culture for participating in organizational citizenship behaviors (OCBs), it is not unreasonable to posit that employees with extremely high agreeableness would engage in high levels of OCBs (e.g., organize birthday celebrations) and employees with extremely low agreeableness would not engage in OCBs (e.g., not attend any birthday celebrations). However, in the middle range of agreeableness scores, the relationship between agreeableness and OCBs would be fairly flat such that for a fairly wide range of agreeableness, there would be no difference in the amount of OCBs displayed.

Trait activation theory (Hochwater, Witt, Treadway & Ferris, 2006; Tett & Burnett, 2003) provides an explanatory mechanism by which personality traits are related to external criteria (e.g., job performance). Trait activation theory posits that individual differences influence outcomes when the situation provides opportunities for trait expression or cues. An individual difference would be expected to positively correlate with job performance when situational cues elicit behavior that contributes to organizational objectives and negatively correlates with job performance when cues elicit behaviors that detract from organizational objectives. When situational cues are

sufficiently salient, most individuals will behave in a similar manner possibly resulting in a cubic relationship between personality and performance.



Figure 4. An illustrative example of a cubic relationship.

It is important to note that the form of nonlinear relationships (e.g., asymptotic, inverted-U, cubic) has important implications for personnel decision-making. For example, an asymptotic model suggests that beyond some threshold, increases in trait levels are not associated with higher performance, whereas an inverted-U model would predict lower performance beyond some specified level of a trait. Unfortunately, some researchers do not explicitly state the hypothesized form of the nonlinear relationship (e.g., Day & Silverman, 1989), and some appear to hypothesize one form (e.g., asymptotic) but test another form (e.g., inverted-U; Le et al., 2010).

In addition to their form (e.g., asymptotic, cubic), nonlinear relationships can also be characterized or described in terms of the number of variables involved in the

relationship. That is, a single variable can display a nonlinear relationship with the criterion when higher-order terms (e.g., squares, cubes) are included in the regression equation. Conversely, nonlinear relationships can occur when two predictor scores interact to predict criterion scores. For example, Barrick and Mount (1991) argued for a single variable asymptotic relationship between emotional stability and performance. Le et al. (2010) proposed a more nuanced multivariable asymptotic relationship where job complexity moderated the relationship between emotional stability and counterproductive work behaviors, such that the point of inflection for highly complex jobs would be at higher levels of emotional stability than the inflection point for less complex jobs (see Figure 5).



Figure 5. An illustrative example of a multiple variable asymptotic relationship.

Personality Variables and Nonlinear Relationships

One advantage of profile matching is its ability to model nonlinear effects. Meta-analytic investigations of personality–performance relationships have generally assumed that these relationships are linear (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000; Tett et al., 1991). However, some researchers have challenged the assumption that performance–personality relationships are linear. Specifically, researchers have argued that conscientiousness (Le et al., 2010), emotional stability (Le et al., 2010), openness (Cucina, & Vasilopoulos, 2005), extraversion (Day & Silverman, 1989), and agreeableness (Graziano, Jenson–Campbell, & Hair, 1996; Graziano & Eisenberg, 1997) may have nonlinear relationships with job relevant outcomes.

For example, conscientiousness appears to be the strongest and most generalizable personality–based predictor of job performance (Barrick & Mount, 1991). However, several researchers have challenged the proposition that conscientiousness is positively and linearly related to job performance (LaHuis, Martin, & Avis, 2005; Le et al., 2010; Robie & Ryan, 1999; Tett, 1998). It is easy to envisage situations where individuals who are either extremely high or extremely low on conscientiousness would perform poorly (Murphy, 1996). Specifically, individuals who are extremely low on this factor may not have the motivation to perform well and tend to behave in a disorganized, irresolute manner. Conversely, individuals who are extremely high on this factor may not perform well in low–structured environments or spend inappropriately large amounts of time on planning and organizing tasks rather than performing their central job functions (i.e., ‘analysis paralysis’; Tett, 1998). Several researchers have found

nonlinear relationships between conscientiousness and several different criteria, including training performance (Vasilopoulos et al., 2007), academic performance (Cucina, & Vasilopoulos, 2005), and job performance (Le et al., 2010).

Robie and Ryan (1999) investigated the possible nonlinear relationship between conscientiousness and job performance, and using 5 samples, failed to find nonlinear relationships between these variables. In contrast, LaHuis et al. (2005) reported nonlinear relationships between conscientiousness and job performance in two samples, with the relationships forming an inverted-U and asymptotic shape for Study 1 and 2, respectively. Le et al. (2010) investigated potential nonlinear relationships between conscientiousness and facets of job performance (i.e., task performance, OCBs, and counterproductive work behavior) using two concurrent designs. The results from Study 1 suggested the presence of a quadratic relationship (i.e., inverted-U) between conscientiousness and all three facets of job performance. However, the Study 2 data failed to replicate these relationships. The authors argue that differences in the operationalizations of personality and performance and characteristics of the samples may explain the differing results. Thus, although conscientiousness displays the largest criterion-related validity coefficient with performance compared to the other personality constructs in the FFM (Barrick & Mount, 1991; Hurtz & Donovan, 2000; cf. Tett et al., 1991), it would seem that there is suggestive evidence that this relationship may not be linear.

In summary, the role of personality in job performance has received considerable research attention and the majority of this research has assumed a linear relationship.

However, in an effort to gain a better understanding of how personality influences job performance, researchers have investigated possible nonlinear relationships and the results are quite inconclusive. For example, four published articles (with a total of 10 samples) have reported nonlinear relationships between conscientiousness (or its facets) and job performance. Two of these articles reported statistically significant nonlinear relationships using three samples (Day & Silverman, 1989; LaHuis et al., 2005). One article reported a failure to find a nonlinear relationship using five samples (Robie & Ryan, 1999), and one article reported a significant nonlinear relationship in one sample but not in the other (Le et al., 2010). Possible explanations for these mixed results include methodological factors (e.g., sample size, personality scale used), potential situational moderators (e.g., job complexity; Le et al., 2010), and possible interactions between personality traits (i.e., trait by trait interactions). As previously noted, profile matching can capture nonlinear effects, and can assess personality in a holistic manner. Indeed, possible explanations for the inconclusive results of nonlinear approaches to personality are potential trait by trait interactions and the tendency for researchers to focus on a single personality factor in isolation.

Whole Person Approaches

The focus on single, narrow personality factors resulted in a negative review of the use of personality in selection by Guion and Gottier (1965). Based on this review, research on personality variables in the context of personnel selection virtually halted until the FFM became a widely accepted taxonomy of personality (Costa & McCrae, 1992; Digman, 1990; cf. Block, 1995; Hough, 1992). The FFM conceptualizes

personality in terms of five relatively broad personality constructs—agreeableness, conscientiousness, emotional stability, extraversion, and openness to experience. Two meta-analyses (i.e., Barrick & Mount, 1991; Tett, et al., 1991) demonstrated the utility of personality tests using the FFM as an organizing framework and concluded that a subset of the FFM personality variables were viable predictors of outcomes of interest. Specifically, Barrick and Mount concluded that conscientiousness was the only FFM variable that demonstrated useful criterion-related validity across a variety of occupational categories (mean $r = .22$; see also Hurtz & Donovan, 2000), whereas Tett et al. concluded that agreeableness displayed the highest level validity coefficient for job performance (mean $r = .33$). The use of FFM predictor constructs to predict broad criterion constructs (e.g., job performance) is reflective of the so-called bandwidth–fidelity dilemma.

The bandwidth–fidelity dilemma (Ones & Viswesvaran, 1996) refers to the choice between using a single narrowly defined and carefully measured trait versus broader traits which are less precise and multifaceted. However, it is important to note that it is possible to construct high fidelity measures of broad personality traits (Ones & Viswesvaran, 1996). Thus, the bandwidth–fidelity hypothesis argues that the similarity of breadth of the predictor and criterion should lead to higher observed relationships. That is, broad traits should better predict broad outcomes, whereas narrow traits should better predict narrow outcomes. Although some researchers argue that information is lost when broad traits are used (e.g., Hough, 1992; Kanfer, Ackerman, Murtha, & Goff, 1995; Oswald & Hough, 2010; Schneider, Hough, & Dunnette, 1996), other researchers

argue for the use of broad personality traits (e.g., Ones & Viswesvaran, 1996). Furthermore, researchers have investigated the efficacy of compound personality traits, including integrity (Berry, Sackett, & Wiemann, 2007; Ones, Viswesvaran, & Schmidt, 1993), customer service orientation (Frei & McDaniel, 1998), core self-evaluations (Erez & Judge, 2001; Judge, Erez, Bono, & Thoresen, 2003), and proactive personality (Crant, 1995).

Since organizationally relevant outcomes are usually quite complex and influenced by multiple individual differences, multiple broad predictors are likely to provide incremental prediction over any single predictor. Thus, the choice of predictors and the method of combining predictors are critical issues. The most common method for combining multiple predictors is to adopt a decision-making strategy that is based on a linear model (e.g., multiple regression, multiple cutoff). For example, it is not uncommon to use all FFM factors to predict outcomes of interest using multiple regression (e.g., Crant, 1995; Judge, Bono, Ilies, & Gerhardt, 2002; Judge et al., 2003; Ones, et al., 2007). However, most research using the FFM has treated these personality factors as independent ignoring possible interactions between them. It is not unreasonable to posit that the expression of one personality variable may depend on another. Two streams of emerging research support this supposition, namely the personality by personality interaction literature and the circumplex models of personality.

Although interactions between personality traits have received some attention in other domains (e.g., occupational health psychology; Zellars, Perrewe, Hochwarter, &

Anderson, 2006), a search of the extant literature resulted in only three studies that examined interactions between FFM factors in predicting job performance. Witt, Burke, Barrick and Mount (2002) first examined the interaction of two FFM factors and its effect on job performance. Specifically, they investigated the interaction between conscientiousness and agreeableness and found that the positive effect of conscientiousness was dependent on the level of agreeableness. However, the interaction effect was only found in jobs that required frequent social interactions. A replication of the Witt et al. study failed to find the same effects (Warr, Batram, & Martin, 2005). A number of methodological differences between the original study and the replication could explain the differences in results. Specifically, Witt et al. used a direct FFM measure, whereas Warr et al. used an indirect measure (a measure that was retrofitted into the FFM framework). Furthermore, Warr et al.'s samples consisted of sales employees, whereas Witt et al.'s samples consisted of sales and nonsales employees.

Judge and Erez (2007) investigated the role of emotional stability and extraversion in job performance. Specifically, they used two operationalizations of the joint influence of emotional stability and extraversion, namely a statistical interaction and a direct measure of the intersection of extraversion and emotional stability based on a FFM circumplex. Their results suggest that both operationalizations provided incremental validity over and above the main effects for emotional stability, extraversion, agreeableness, and conscientiousness. So, although interactions between

FFM personality factors have received relatively little empirical attention, the results from these studies suggest this is a potentially fruitful area for future research.

Circumplex models (De Raad, Hendricks, & Hofstee, 1994; Hofstee, De Raad, & Goldberg, 1992; Johnson & Ostendorf, 1993) represent a divergence from the simple structure perspective of personality commonly employed in personality research. Specifically, Hofstee et al. (1992) present data that suggest a large proportion of trait variables load on two factors and approximately half of the trait variables loaded on only one factor suggesting that a simple structure perspective may not represent personality data very well. Thus, De Raad and Hofstee (1993) advocate for the use of the Abridged Big-Five Circumplex Model (AB5C) which consists of ten bidimensional circumplex models which represent all pairs of the FFM dimensions (see Figure 6). An examination of the AB5C suggests that trait terms are not evenly distributed across the FFM dimensions. This implies, for example, that there are several trait descriptors for a person who is both high on agreeableness and high on extraversion, but few or no trait descriptors for a person low on extraversion and high on agreeableness.

Although profile matching provides a number of advantages over linear models, there is a dearth of research investigating the efficacy of profile matching for personnel decision-making. A search of the extant literature found no published studies, one unpublished dissertation (Jenkins, 2002), and one conference paper (Waters & Sackett, 2006) that compared profile matching with linear regression models in terms of personality and performance. Although profile matching has not received much attention in the I/O psychology personality literature, it has received some attention in

predicting student success in college (e.g., Mumford & Owens, 1982; Schmitt, Oswald, Kim, Imus, Merritt, Friede, & Shivpuri, 2007), vocational interest (e.g., Dilchert, 2007;

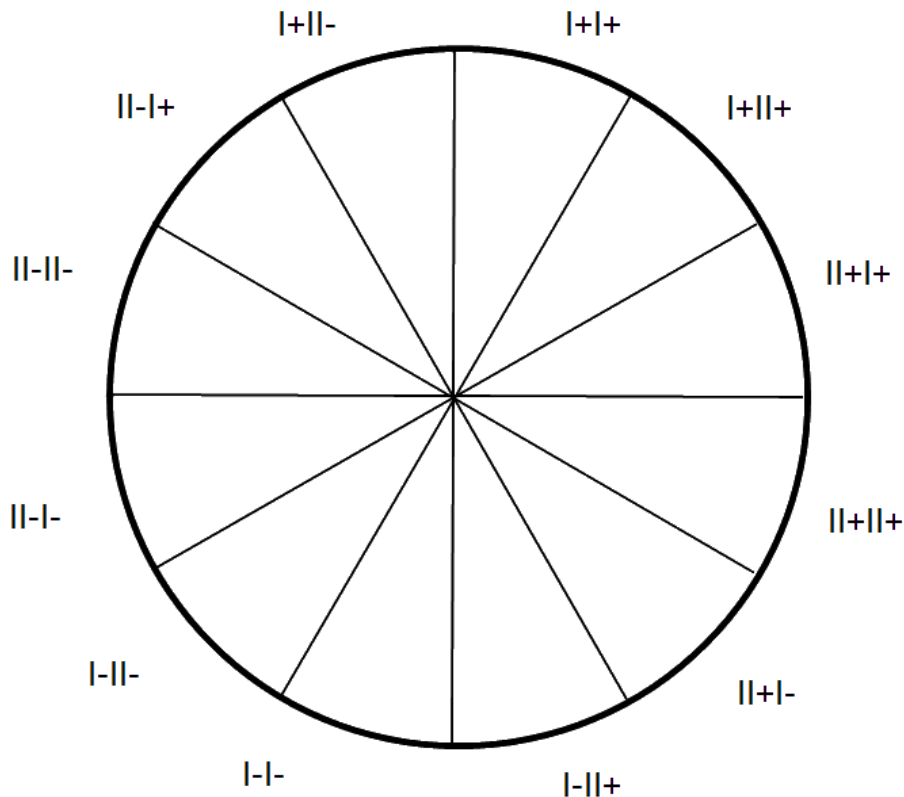


Figure 6. An illustrative example of a circumplex for extraversion (I) and agreeableness (II). Adapted from DeRaad et al. (1994, p. 95).

Holland, 1985; Owens & Schoenfeldt, 1979), and diagnosing personality disorders (e.g., Hathaway & Meehl, 1951; Lynam & Widiger, 2001). In these domains, profile matching generally outperforms linear composites in predicting the outcomes of interest. For example, Davison and Davenport (2002) examined the relationship between personality and vocational interest. They found that a profile matching approach explained 12.7% of the variance over and above that explained by a linear composite.

Similarly, Dilchert (2007) investigated the relationship between vocational interest and personality in terms of personality profile level and pattern. These results suggest that the profile pattern was a better predictor of interest in leadership and supervision compared to profile level. For example, profile pattern was associated with more variance in interest for leadership ($\Delta R^2 = .15$) compared to that of the profile level ($\Delta R^2 = .02$) when personality was measured via the NEO PI-R. Schmitt et al. (2007) used a clustering method to assess the extent to which background and ability profiles predicted college student outcomes. Their analysis resulted in five distinct predictor profiles. However, unlike Davison and Devenport (2002), and Dilchert (2007), Schmitt et al. (2007) found that profile configuration did not display incremental validity over a linear composite. Specifically, the linear composite of the predictors resulted in multiple correlations ranging from .18 to .72 for various outcomes (e.g., first year grade point average, intent to quit college). The change in the multiple correlations associated with the inclusion of the profile analysis ranged from .001 to .005, and none of these changes were statistically significant.

In summary, profile matching has not received much attention in the I/O psychology literature in general, and the I/O psychology personality literature in particular as a personnel decision-making strategy. The use of linear models may exacerbate problems associated with response distortion. The reliance on linear models may be due to the challenges associated with nonlinear research. However, profile matching has a number of potential benefits, including mitigating response distortion. That is, since profile matching strategies capture the *configuration* of predictor scores, it

is posited that effectively distorting one's response would be rather difficult since the most desirable score is not necessarily at an extreme end of the scale. Furthermore, the desired level of a specific trait depends on the levels of the other traits contained in the profile.

It is important to note that the focus of the current study is on the extent to which profile matching may alleviate concerns regarding response distortion. The study objectives were accomplished by comparing profile fit scores and linear composite scores in terms of the extent to which they correlate with lie scale scores and turnover, and the extent to which the criterion-related validities are affected by corrections for response distortion. If profile matching and linear composite scores are found to be similar in terms of their criterion-related validities and profile fit scores are less susceptible to response distortion, then profile matching may be preferred if response distortion is an overriding concern. That is, for profile matching to be a viable response distortion reduction strategy, profile fit scores must display similar or better criterion-related validity compared to a linear composite. Furthermore, if response distortion attenuates the criterion-related validity of personality tests scores which are based on linear composites, then profile matching may provide gains in criterion-related validity.

CHAPTER IV

CURRENT STUDY

The primary objective of the current study was to investigate and document the extent to which concerns about response distortion may be alleviated by use of profile matching strategies in personnel selection decision-making. A secondary objective was to investigate the extent to which the configuration of the standard profile influences the efficacy of profile matching as a means for reducing response distortion. For the purposes of the current investigation, profile matching fit scores were compared to linear composite scores in terms of their susceptibility to response distortion, observed criterion-related validities in predicting turnover (tenure and voluntary turnover), and levels of subgroup differences.

Susceptibility to Response Distortion

Effectively distorting one's responses would seem to be relatively easier under a decision-making strategy based on a linear model compared to a profile matching strategy. Specifically, engaging in response distortion when a linear composite score is used would entail determining the valence of the trait (i.e., if the trait described was desirable or undesirable) and responding in an overly extreme manner accordingly. The efficacy of extreme responding would appear to be a function of the transparency (i.e., the extent to which the test-taker can identify the construct measured) and the valence (i.e., which response option is considered the most desirable from the test-user's perspective) of test items.

Conversely, under a profile matching strategy, the primary mechanism that mitigates response distortion is the configural structure of the ideal profile. That is, the ideal profile is determined by the pattern of scores across multiple personality variables. Thus, effectively distorting one's responses would be a function of the extent to which test-takers know the ideal configuration, and, because the ideal configuration is going to vary across situations and jobs, the likelihood that test-takers will know the ideal pattern is quite low. It would seem that the number of variables contained in the profile would be inversely related to the likelihood that applicants could know or guess the ideal profile. Even for stereotypical jobs, it would be difficult to guess the ideal configuration if the number of variables contained in the profile was large. As such, the use of only extreme responses would be an effective response distortion strategy under the linear model, but would not be effective under a profile matching strategy.

To investigate and document the extent to which response distortion concerns may be alleviated by using a profile matching strategy, profile fit scores and linear composite scores were computed using a sample of 996 applicants completing an operational personality test. Missing data were imputed for a subset ($n = 291$) of applicants who did not complete the response distortion scales. As such, the present study represents an initial test of the viability of profile matching as a means for alleviating concerns about response distortion using a blend of real and simulated data. The proportion of imputations is what necessitates the need to interpret the current study more as a simulation, and thus, as a proof-of-concept, instead of a field test of the ideas presented here.

Because the profile fit score is based on a configural structure of the predictor scores, whereas the linear composite is not, these two scores should represent different manifestations of the personality domain space. As such, it was expected that the correlation between the tenure profile fit scores (i.e., fit scores which represent the difference between the test-taker's personality profile and that of a group test-takers whose tenure was greater than one year) and the linear composite scores would be modest. Thus, it was hypothesized that:

Hypothesis 1: The observed correlation between the tenure profile fit score and linear composite will be modest ($r < .50$).

To investigate the extent to which profile fit scores and linear composite scores are influenced by response distortion, the association between these scores and various lie scale scores was assessed. As previously mentioned, several types of lie scales which vary in the types of response distortion they detect are used in research and practice. Specifically, the improbable item approach uses items that are highly desirable but unlikely to be true. Test-takers who endorse a large number of these items at high levels of agreement are suspected of distorting their responses. Another approach to detecting response distortion entails the use of lie scales based on consistency of responses. These lie scales posit that response distortion is associated with lower levels of consistent responses. That is, this method of response distortion detection is predicated on the supposition that test-takers who distort their responses will not respond consistently to item pairs, where one item is socially desirable and the other item is socially undesirable, whereas an honest test-taker would respond consistently (Hand, 1964).

High scores on the improbable item and consistency-in-responding lie scales are considered indicators of a test-taker's propensity to engage in response distortion. Thus, relationships between these lie scale scores and the focal personality scores would suggest that test-takers who distort their responses on the lie scales also scored high on the external measure and that the external measure is susceptible to response distortion. Based on the supposition that profile fit scores should be comparatively more resistant to response distortion than the linear composite scores, it was expected that the profile fit scores would display smaller correlations with the lie scale scores compared to that of the linear composite scores. Specifically, it was hypothesized that:

Hypothesis 2: The observed correlation between tenure profile fit scores and (a) the improbable item lie scale scores, and (b) the consistency-in-responding lie scale scores will be smaller than that of the linear composite scores and the corresponding lie scale scores.

Criterion-Related Validity

Employee tenure is an important outcome to organizations as high rates of turnover have a number of negative consequences including loss in productivity (Huselid, 1995), and costs associated with restaffing. Many antecedents to turnover have received considerable research attention, including the work environment (Griffeth, Hom, & Gaertner, 2000), attitudes (Johns, 2002), and personality (Zimmerman, 2008). With regard to personality, Barrick and Mount's (1991) meta-analytic results suggest a weak relationship between the FFM personality variables and turnover/tenure. Specifically, they found sample-

weighted mean correlations of -.03, .01, .06, .09, and -.08 for extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience, respectively (see also Salgado, 2002). However, they did not differentiate between voluntary and involuntary turnover. Studies that have differentiated between voluntary and involuntary turnover have generally found larger criterion-related validities (e.g., Barrick & Mount, 1996; Barrick, Mount, & Strauss, 1994). For example, Zimmerman (2008) examined the relationship between the FFM factors and voluntary turnover. Zimmerman reported meta-analytic estimates of -.03, -.16, -.22, -.18, and .09 for extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience, respectively.

One potential explanation for these relatively small criterion-related validities is that response distortion may be attenuating these relationships. Indeed, researchers and practitioners continue to be concerned about the effects of response distortion (e.g., Tett & Christiansen, 2007). Thus, although the effects of response distortion on the criterion-related validities of personality test scores have received considerable research attention, as previously noted, there continues to be debate about the extent to which response distortion attenuates these relationships (e.g., Morgeson et al., 2007; Ones et al., 2007; Tett & Christiansen, 2007). As such, an objective of the current study is to compare profile fit scores to linear composite scores in terms of their ability to predict turnover. If response distortion poses a threat to the criterion-related validity of personality test scores, then it is reasonable to expect that the correlation between a

linear composite score and tenure/turnover (i.e., tenure and voluntary turnover) would be smaller than that for the profile fit score. As such, the following hypotheses were tested:

Hypothesis 3a: The criterion-related validity of the tenure profile fit scores in predicting tenure will be larger than that of the linear composite scores.

Hypothesis 3b: The criterion-related validity of the voluntary turnover profile fit scores in predicting voluntary turnover will be larger than that of the linear composite scores.

It is acknowledged that the hypothesized differences between the criterion-related validities of the profile fit scores and the linear composite scores could be due to either differences in their susceptibility to response distortion or differences in the manner in which the relationship is modeled, or both. That is, either response distortion, model misspecification, or both could lead to profile fit scores demonstrating larger criterion-related validities compared to the linear composite scores. To examine the extent to which profile fit scores are less susceptible to response distortion compared to the linear composite scores, the criterion-related validities of the profile fit scores and linear composite scores were corrected for response distortion. If the profile fit scores are resistant to response distortion, then correcting for response distortion should not result in practically significant differences between the corrected and uncorrected validities. Conversely, if the linear composite score is susceptible to response distortion, then correcting for response distortion should result in statistically and

practically significant differences between the corrected and uncorrected validities.

Hypothesis 4: The differences in corrected and uncorrected criterion-related validity of the tenure profile fit scores in predicting tenure will be minimal when profile fit scores are corrected on the basis of (a) the improbable item lie scale scores and (b) the consistency-in-responding lie scale scores.

Hypothesis 5: The differences in corrected and uncorrected criterion-related validity of the linear composite scores in predicting tenure will be significant when linear composite scores are corrected on the basis of (a) the improbable item lie scale scores and (b) the consistency-in-responding lie scale scores.

Hypothesis 6: The differences in corrected and uncorrected criterion-related validity will be larger for the linear composite score compared to the profile fit score when these scores are corrected on the basis of (a) the improbable item lie scale scores and (b) the consistency-in-responding lie scale scores.

Subgroup Differences

One touted advantage of using noncognitive predictors in general, and personality variables in particular, is the relatively small observed subgroup differences compared to other common job selection constructs (e.g., Foldes, Duehr, & Ones, 2008; Ones & Anderson, 2002; Ployhart & Holtz, 2008). Using predictors that display small

or negligible subgroup differences reduces the likelihood that selection decisions will display adverse impact against protected classes as subgroup differences are a necessary but not sufficient condition of adverse impact.

Several meta-analyses have investigated subgroup differences across a wide range of predictors (e.g., Hough, Oswald, & Ployhart, 2001; Ployhart & Holtz, 2008). The results from these meta-analyses generally support the claim that personality variables tend to display small to negligible race-based subgroup differences. However, meta-analytic evidence suggests there are substantial sex-based subgroup differences for two of the FFM dimensions—agreeableness and emotional stability. Specifically, Ployhart and Holtz report sample-weighted standardized mean differences of -0.39 and 0.24 for agreeableness and emotional stability, respectively (using males as the comparator such that a positive value indicates males score higher than females). Extraversion, conscientiousness, and openness to experience tend to display negligible sex-based subgroup differences. As such, the claim that the use of personality variables as predictors is an effective strategy to reduce adverse impact should be qualified by the specific personality trait and protected class variables of interest. Additional qualifications regarding personality subgroup differences are required when using facets of the FFM as opposed to the factors. For example, although emotional stability displays quite small differences between African-Americans and White subgroups ($d = -0.12$), facets of emotional stability display markedly varied levels of subgroup difference ($d = -0.23$ to 0.17 ; Foldes et al., 2008).

The purpose of examining the extent to which profile fit scores and linear composite scores display different patterns of subgroup differences is two-fold. First, the expectation is that profile fit scores and linear composite scores would display different manifestations of the personality domain space. Thus, different patterns of subgroup differences would further support the supposition that profile fit scores and linear composite scores represent different conceptualizations or manifestations of the domain space. Second, the extent of subgroup differences is a critical outcome for any predictor-method, predictor-construct, or personnel decision-making strategy. That is, from a practitioner perspective, it is common to evaluate selection devices in terms of their validity, utility, and extent of subgroup differences. Given the relative lack of research regarding personality profile matching as a personnel decision-making approach, there is no a priori reason to expect profile matching to result in lower or higher levels of subgroup difference. However, given the importance of subgroup differences in personnel research and practice (Foldes et al., 2008; McDaniel, Kepes, & Banks, 2011; Ones & Anderson, 2002), it is critical to document the relative levels of subgroup differences resulting from the two methods of generating combined scores.

Profile Configuration

The efficacy of profile matching strategies in reducing response distortion is likely a function of the configuration of the standard profile used. Indeed, a linear composite can be conceptualized as a standard profile where all components take on extreme values as the ideal score. Using a profile matching strategy would not be expected to reduce response distortion if the standard profile had uniformly high ideal

scores. Conversely, standard profiles that take on moderate ideal values for a subset of the profile components would be expected to mitigate response distortion to the extent that the ideal values were not known. For example, Figure 7 presents three hypothetical profiles.

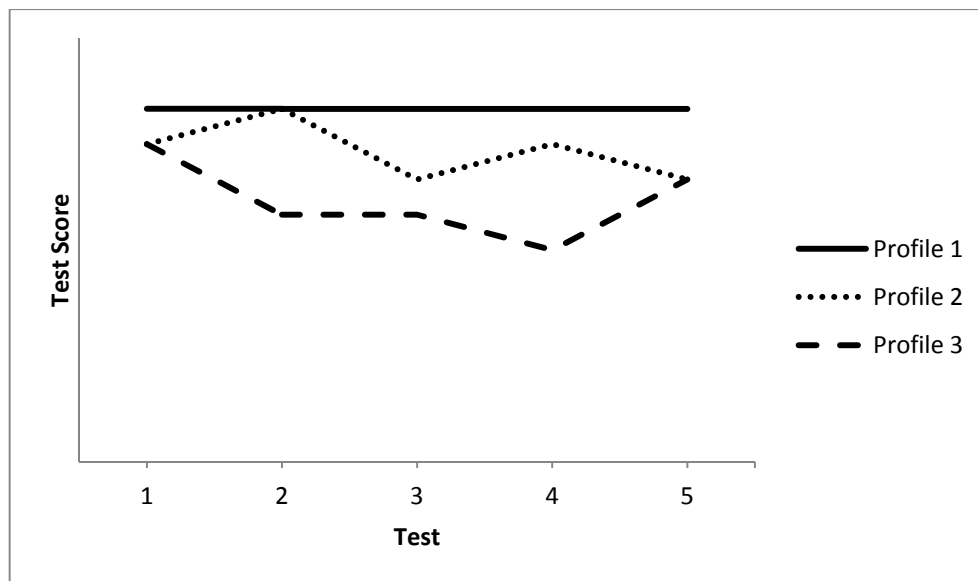


Figure 7. Three example profiles which differ in terms of profile complexity.

Test score profiles can be described in terms of their elevation and scatter (or dispersion; Cronbach, & Gleser, 1953; Nunnally & Bernstein, 1994). The elevation of the standard profile represents the magnitude of the ideal trait levels across the profile. Elevation is operationalized as the mean of all scores in the profile. Scatter represents the extent to which the scores in a profile vary from one another. Scatter is the square root of the sum of squared deviations about the mean score (i.e., its elevation).

Figure 7 presents three profiles that differ in terms of their elevation and scatter and Table 1 presents the component scores, elevation, and scatter. Profile 1 represents a test score profile with the maximum elevation and no scatter (i.e., a linear composite). Using a profile matching strategy would not be expected to reduce response distortion if the standard profile had uniformly high ideal scores as represented by Profile 1. Conversely, Profiles 2 and 3 would be expected to mitigate response distortion when test-takers are unaware of the standard profile's configuration. That is, standard profiles that have moderate elevation and higher degrees of scatter are posited to alleviate the effects of response distortion to a greater extent than standard profiles that have higher (or lower) elevation and lower degrees of scatter.

Cronbach and Gleser (1953) present the following equation as a method of assessing the difference between two profiles:

$$D_{12}^2 = \sum_{j=1}^k (x_{j1} - x_{j2})^2 \quad [\text{EQ 1}]$$

where D = distance, j = the profile components, k = number of profile components, i = profile, and x_{ij} = the score for profile i on component j . It is important to note that the distance score reflects misfit, such that smaller distance scores represent higher profile fit. The distance score can be decomposed into three components, namely elevation, scatter, and shape. Removing the effects of elevation from the distance score is accomplished by mean-centering the test scores within each profile.

The distance between profiles after controlling for elevation is represented by the following equation:

$$D'_{12}{}^2 = D_{12}^2 - k\Delta^2 El_{12} \quad [\text{EQ } 2]$$

where D' = distance controlled for elevation, D = distance, k = number of profile components, $\Delta^2 El_{12}$ = difference in elevation between the two profiles. Removing the

Table 1

Illustrative Profile Component Scores, Elevation, and Scatter

Profile	Test					Elevation	Scatter
	1	2	3	4	5		
1	5.0	5.0	5.0	5.0	5.0	5.00	0.00
2	4.5	5.0	4.0	4.5	4.0	4.40	0.84
3	4.5	3.5	3.5	3.0	4.0	3.70	1.14

effects of elevation and scatter from the distance score can be accomplished by mean-centering the test scores within each profile, and then dividing by the profile scatter. The distance between profiles due to shape is represented by the following equation:

$$D''^2 = \frac{D'^2 - \Delta^2 S}{S_1 S_2} \quad [\text{EQ } 3]$$

where D'' = distance controlled for elevation and scatter, D' = distance controlled for elevation, $\Delta^2 S$ is the difference in scatter; S_1 = scatter for Profile 1; S_2 = scatter for Profile 2.

Applying Equation 1 to the illustrative profiles (see Figure 7 and Table 1) produces a distance score of 2.50, 9.75, and 4.75 for the difference between Profile 1 and Profile 2, Profile 1 and Profile 3, and Profile 2 and Profile 3, respectively. Applying Equation 2 results in a distance score which is controlled for elevation of 0.70, 1.30, and

2.30 for the difference between Profile 1 and Profile 2, Profile 1 and Profile 3, and Profile 2 and Profile 3, respectively. Finally, Equation 3 results in a difference in shape of 2.31 between Profile 2 and Profile 3. Because Profile 1 has no scatter, differences in shape could not be calculated for comparisons between Profile 1 and the other illustrative standard profiles. The use of Equations 1–3 allow for a quantitative description of the extent to which two standard profiles differ.

To examine the extent to which the number of components that do not take on extreme ideal values influences the efficacy of a profile matching strategy, three simulated ideal profiles were used. Specifically, three profiles that have no scatter, moderate levels of scatter, and high levels of scatter were generated.

The efficacy of profile matching would appear to be contingent on the configural nature of the standard profile. That is, it is posited that profile matching will mitigate response distortion when test-takers do not know the ideal trait levels and standard profiles that have varying ideal trait levels across profile components (i.e., standard profiles with a larger distance score from a uniformly high standard profile) will be less likely to be known or guessed by the test-takers compared to a standard profile that has less variation in ideal trait levels. Hence, it was hypothesized that:

Hypothesis 7. The profile fit scores based on an ideal profile with no scatter will display a larger correlation with (a) the improbable item lie scale scores and (b) the consistency-in-responding lie scale scores compared to the profile fit scores based on a profile with moderate levels of scatter.

Hypothesis 8: The profile fit scores based on an ideal profile with moderate levels of scatter will display larger correlations with (a) the improbable item lie scale scores and (b) the consistency-in-responding lie scale scores compared to the profile fit scores based on a profile with high levels of scatter.

If profile complexity is related to susceptibility to response distortion, then it is expected that these differences would be manifested in the observed criterion-related validity of profile fit scores. It was expected that response distortion would attenuate the criterion-related validity of both profile fit scores based on an ideal profile with no scatter and, to a lesser extent, the profile fit scores based on an ideal profile with moderate levels of scatter. However, it was expected that the profile fit scores based on an ideal profile with high levels of scatter would be relatively more resilient to response distortion. As such, it was hypothesized that:

Hypothesis 9a: The profile fit scores based on a standard profile with high levels of scatter will display a larger criterion-related validity coefficient compared to the profile fit scores based on a standard profile with moderate levels of scatter.

Hypothesis 9b: The profile fit scores based on a standard profile with moderate scatter will display a larger criterion-related validity coefficient compared to the profile fit scores based on a standard profile with no scatter.

To examine the extent to which differences in the criterion–related validities of these three scores are due to response distortion, they were corrected on the basis of lie scale scores. Since it was expected that the differences in the criterion–related validities of these three test scores are due to their susceptibility to response distortion, it was hypothesized that:

Hypothesis 10: The profile fit scores based on a standard profile with high levels of scatter will display a minimal difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on (a) the improbable item lie scale, and (b) the consistency-in-responding lie scale scores.

Hypothesis 11: The profile fit scores based on a standard profile with moderate levels of scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on (a) the improbable item lie scale, and (b) the consistency-in-responding lie scale scores.

Hypothesis 12: The profile fit scores based on a standard profile with no scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on (a) the improbable item lie scale, and (b) the consistency-in-responding lie scale scores.

Hypothesis 13: The differences between corrected and uncorrected criterion-related validities will be smaller for profile fit scores based on

the standard profile with high levels of scatter compared to that of the profile fit scores based on a standard profile with moderate levels of scatter when corrections are based on (a) the improbable item lie scale, and (b) the consistency-in-responding lie scale scores.

Hypothesis 14: The difference between corrected and uncorrected criterion-related validities will be smaller for the profile fit scores based on a standard profile with moderate levels of scatter compared to that of the profile fit scores based on a standard profile with no scatter when corrections are based on (a) the improbable item lie scale, and (b) the consistency-in-responding lie scale scores.

Assessing Fit

The assessment of fit between two sets of scores is of interest in several domains in psychology (e.g., clinical, counseling, I/O). A critical issue in fit research is the statistical method used to assess fit. Several researchers have noted that results of various studies vary as a function of the fit statistic used, and they have also discussed the advantages and disadvantages of using various statistical approaches (e.g., Cronbach & Gleser, 1953; Edwards, 1994). Fit has often been operationalized using difference scores (see Edwards, 1994). However, difference scores engender a number of disadvantages including the potential for low reliability, biased parameter estimates, and restricted variance (Edwards, 2001a). Edwards and colleagues (e.g., Edwards, 2001b; Edwards & Cooper, 1990; Edwards & Parry, 1993) have documented the advantages and appropriate use of polynomial regression as the preferred alternative to the use of

difference scores in assessing fit. Polynomial regression is the appropriate statistical approach to the study of person–organization (P–O) fit and similar constructs (e.g., person–environment fit). Polynomial regression simultaneously assesses fit between two sets of variables, where those two sets of variables vary across observations. Using P–O fit as an example, variables associated with the person and variables associated with the organization are expected to vary for each person. In the present case, the standard profile is constant across test–taker. As such, polynomial regression is inappropriate as every regression term associated with the standard profile will, by definition, be zero. As such, the current study used the difference score described by Cronbach and Gleser (1953) and the results were interpreted within the context of the limitations engendered by its use. Specifically, Equation 1 was used to assess fit between the standard profiles and the test–takers’ profiles.

CHAPTER V

METHOD

Participants

The sample consisted of 996 applicants who completed a personality test as part of an operational unproctored internet-based selection testing assessment for a national retail chain. Only data for applicants who were subsequently hired were available in the archival dataset. Hiring decisions were made partly on the basis of personality scores using a proprietary profile fit score (i.e., profile fit scores that are different from the profile fit scores used in the current study). The implications for range restriction are discussed in the Discussion section. The data were collected between March of 2007 and January 2010 by a large internet-based testing firm, which tests approximately 20,000 test-takers per week. The mean age of the applicants was 25.99 years ($SD = 8.72$; 3.11% of the sample did not report their age). There were 566 females and 425 males with five applicants not reporting their sex. There were 209 African-Americans, 7 American Indians, 18 Asians, 6 Pacific Islanders, 501 Hispanics, and 220 Caucasians, with 18 applicants reporting their race as 'other' and 17 applicants not reporting their race.

Measures

Personality. Applicants completed one of two unproctored internet-based untimed personality tests. Two hundred ninety-one applicants completed the first personality test which consisted of the Guilford-Zimmerman Temperament Survey (Guilford, Zimmerman, & Guilford, 1976) and the Differential Personality Inventory

(Jackson & Messick, 1964). The Guilford–Zimmerman consisted of nine scales and a total of 110 items and the Differential Personality Inventory, eight scales and 95 items. These scales were previously sorted by Arthur et al. (2009, 2010) into the FFM dimensions using the guidelines and procedures presented in Barrick and Mount (1991) and Birkeland et al. (2006). Arthur et al. (2010) reported score internal consistency reliability coefficients of .70, .73, .72, .81, and .68 for agreeableness, conscientiousness, emotional stability, extraversion, and openness, respectively. Because the dataset only contained scale–level data, score internal consistency reliability coefficients could not be calculated for the current dataset.

The remaining applicants ($n = 705$) completed a proprietary internet–based untimed test (see Arthur, Villado, & Glaze, 2007) that consisted of 22 10–item personality scale items and a 10–item improbable item scale for a total of 230 items. Applicants responded to test items using a 5–point Likert scale. These scales were previously sorted into the FFM dimensions by Arthur et al. (2009, 2010). Arthur et al. (2010) reported score internal consistency reliability coefficients of .81, .95, .85, .93, .81, and .84 for agreeableness, conscientiousness, emotional stability, extraversion, openness, and the improbable item lie scale, respectively.

Improbable Item Lie Scale. Response distortion was operationalized using two lie scales. First, the improbable item lie scale consisted of 10 improbable items. Improbable item lie scale scores were computed as the number of times a test–taker endorsed an improbable item with “Strongly Agree”. Improbable item lie scale scores were not available for 291 applicants. Consequently, a simple regression imputation

strategy (Roth, 1994) was used to impute the missing data. Specifically, improbable item lie scale scores were regressed on to the five FFM personality dimensions using records with complete data. The predicted improbable item lie scale score was used to replace missing values.

Consistency–In–Responding Lie Scale. The consistency of responding lie scale consisted of 10 personality test item pairs (with one positive item and one negative item per pair) which one would expect test-takers to respond to in a consistent manner. Consistency–in–responding lie scale scores were computed as the absolute difference in responses on the pairs of consistent items such that larger scores reflect higher levels of response distortion. Improbable item lie scale scores were not available for 291 applicants. So again, a simple regression imputation strategy (Roth, 1994) was used to impute the missing data.

Tenure. Voluntary and involuntary turnover data were gathered from company records approximately one year after initial testing. The majority (77.01%) of test-takers remained in the organization after one year. Managers of the local retail chains documented the reason for the employee departures, and the testing firm subsequently categorized each departure as voluntary or involuntary. In addition, the dataset contained the hire date for all employees along with the termination date for employees who were terminated or quit when this information was collected. The tenure data consisted of the number of days an employee worked before turning over. Employees who had not turned over when the criterion data were collected received a tenure value as if they turned over the day after the criterion data were collected.

Profile Fit Scores. Two standard profiles were generated for the purposes of the current study. To generate the standard profile scores for predicting voluntary turnover, test-takers who turned over voluntarily within one month were identified, and a random subset of 20 (25.64%) of those applicants identified were used as the reference group. The standard profile was the mean of those applicants' scores on all the FFM dimensions. Table 2 presents the means and standard deviations used for the voluntary turnover standard profile. Table 3 presents the voluntary turnover standard profile with its profile characteristics (i.e., elevation and scatter). This process was repeated to generate the other standard profile (i.e., tenure standard profile). Specifically, applicants who remained with the organization longer than one year were identified, and a random subset of 20 (10.87%) of those applicants identified were used as the reference group. The standard profile for tenure was the mean of those applicants' scores on all of the FFM dimensions.

The test-takers in the reference groups were excluded from the rest of the analyses (i.e., sampling without replacement). Fit scores were generated between the two standard profiles and test-taker profiles using Equation 1. Because distance scores are used, higher profile fit scores represent greater degrees of misfit compared to lower profile fit scores.

Simulated Profile Fit Scores. An examination of Table 3 suggests that the tenure standard profile and the voluntary turnover standard profile do not meaningfully differ in scatter. To examine the influence of scatter on the efficacy of profile matching in reducing the effects of response distortion, profiles with different levels of scatter are

Table 2

Means, Standard Deviations, and Sample Sizes for the Two Referent Groups and the Overall Sample

Variable	<i>N</i>	Mean	<i>SD</i>
Voluntary Turnover Reference Group			
Agreeableness	20	54.80	15.78
Conscientiousness	20	52.32	16.92
Emotional Stability	20	54.98	14.54
Extraversion	20	52.56	13.21
Openness	20	46.33	9.50
Tenure Reference Group			
Agreeableness	20	57.27	13.80
Conscientiousness	20	54.30	7.95
Emotional Stability	20	59.74	11.27
Extraversion	20	54.78	8.98
Openness	20	45.43	7.58
Overall Sample			
Agreeableness	956	54.42	16.40
Conscientiousness	956	54.14	14.88
Emotional Stability	956	55.57	14.90
Extraversion	956	52.09	10.39
Openness	956	43.78	8.68

required. Since the tenure and voluntary turnover standard profiles do not meet these requisite conditions, simulated profiles were used. Thus, three profiles were generated which differed in terms of profile scatter, but not in profile elevation. The simulated standard profiles are presented in Table 4. Profile 1 has no scatter, Profile 2 has a moderate amount of scatter, and Profile 3 has a high level of scatter.

Table 3

Standard Profiles for Voluntary Turnover and Tenure

Profile	Personality Dimension					Profile Characteristics	
	Agree	Consc	Emot	Extr	Open	Elevation	Scatter
Voluntary Turnover	54.80	52.32	54.98	52.56	46.33	52.20	7.01
Tenure	57.27	54.30	59.74	54.78	45.43	54.30	10.83

Note. Agree = agreeableness; Consc = conscientiousness; Emot = emotional stability; Extr = extraversion; Open = openness.

Linear Composite Scores. A unit-weighted linear composite was computed for each test-taker. The linear composite contained all FFM dimensions. The unit-weighted composite scores were computed as the average of the scale scores contained in the composite. The decision to compare profile fit scores to linear composite scores instead of a regression-based approach warrants some discussion. The primary reason for using a linear composite (i.e., unit-weighted composite) score was that this was deemed to be the most direct analog to a (unit-weighted) profile fit score. Although it is possible to develop differentially-weighted profile fit scores, the development of these fit scores is beyond the scope of this dissertation. Furthermore, the use of a regression-based linear composite capitalizes on chance, and thus, comparing a unit-weighted profile fit score to a regression-based linear composite would unduly bias the results in favor of the linear composite. Based on these considerations, a unit-weighted linear composite was used.

Table 4

Simulated Standard Profiles

Simulated Profile	Personality Dimension					Profile Characteristics	
	Agree	Consc	Emot	Extr	Open	Elevation	Scatter
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	-0.25	0.00	0.50	0.00	-0.25	0.00	0.61
3	-0.50	0.00	1.00	0.00	-0.50	0.00	1.22

Note. Agree = agreeableness; Consc = conscientiousness; Emot = emotional stability; Extr = extraversion; Open = openness.

Procedure

Applicants completed the personality test under unproctored internet-based testing conditions. Demographic information and turnover data were provided by the testing firm. Turnover data were collected approximately one year after testing.

CHAPTER VI

RESULTS

Empirical Profile Fit Scores

Relationship Between Profile Fit and Linear Composite Scores. It was posited that profile fit and linear composite scores would represent different manifestations of the personality domain space. That is, if profile fit scores and linear composite scores represent the same manifestation of personality, one would expect correlations that are more commensurate with test–retest or equivalent forms reliability which are expected to have fairly high correlations (e.g., .80 or higher). Hypothesis 1 stated that the correlation between the tenure profile fit scores and linear composite scores would be modest ($r < .50$). The relationship between tenure profile fit scores and linear composite scores was small ($r = .09, p < .05$) and significantly smaller than .50 ($z_r = -14.17, p < .05$). Thus, Hypothesis 1 was supported. Although not specifically hypothesized, the voluntary turnover profile fit score also demonstrated a small correlation with the linear composite score ($r = .21$), which is significantly smaller than $r = .50$ ($z_r = -10.38, p < .05$). The Simulated Profile Fit scores displayed correlations that ranged from .25 to .35 with the linear composite scores. All of the Simulated Profile Fit scores demonstrated correlations that were statistically smaller than .50.

Susceptibility to Response Distortion. To investigate the differential susceptibility to response distortion between profile fit scores and linear composite scores, the equality of correlations between the specified lie scale scores and the tenure profile fit and linear composite scores was evaluated. The correlation between the

improbable item lie scale and the tenure profile fit score was positive and moderate ($r = .22, p < .05$) such that test-takers who distorted their responses tended to have worse fit scores (high profile fit scores indicate poor fit). The relationship between the improbable item lie scale and the linear composite score was large and positive ($r = .73, p < .05$) indicating test-takers who engaged in response distortion had more favorable linear composite scores. The difference between these two correlations ($r_d = -.51$) was statistically significant ($t [953] = -16.05, p < .05$), supporting Hypothesis 2a. The relationship between the consistency-in-responding and the profile fit scores ($r = .05, p > .05$) was small and indicated that test-takers who engaged in response distortion had worse fit scores compared to test-takers who responded honestly. The consistency-in-responding lie scale scores displayed a negative correlation with the linear composite scores ($r = -.09, p < .05$), indicating that test-takers who distorted their responses tended to score lower on the linear composite. This relationship was in the opposite direction than expected. As such, the difference in correlation cannot be meaningfully interpreted within the framework of the hypothesis. However, for the sake of completeness, the difference ($d_r = 0.14$) was statistically significant, $t (953) = 3.22, p < .05$.

Criterion-Related Validity. Hypothesis 3a posited that the tenure profile fit score would display a stronger criterion-related validity compared to the linear composite in predicting tenure. The tenure profile fit score displayed a negative relationship with tenure ($r = -.16, p < .05$), which indicates that test-takers with better fit had longer tenure. This relationship was stronger in magnitude compared to the relationship between the linear composite and tenure ($r = .07, p < .05$), which also indicated that test-

takers with higher linear composite scores had longer tenure. The difference between the absolute values of the coefficients was statistically significant, $t(953) = 2.08, p < .05$. Thus, Hypothesis 3a was supported. The relationship between the voluntary turnover profile fit scores and voluntary turnover was small and negative ($r = -.06, p > .05$). This suggests that test-takers with better fit scores were more likely to voluntarily turnover compared to test-takers with worse fit scores. The relationship between the linear composite score and voluntary turnover was small ($r = 0.07, p < .05$). Thus, test-takers with higher linear composite scores were more likely to voluntarily turnover. The difference between the absolute values of the correlations was not statistically significant, $t[953] = -0.25, p > .05$. Thus, Hypothesis 3b, which stated that the criterion-related validity of the voluntary turnover profile fit score would display a stronger relationship with voluntary turnover compared to the linear composite score, was not supported.

To investigate the effects of correcting for response distortion on the criterion-related validity of the tenure profile fit scores, the semipartial correlation between the tenure profile fit scores and tenure was computed, partialing out the improbable item lie scale score from the tenure profile fit scores. The correlation between tenure profile fit scores and tenure ($r = -.16$) was smaller in magnitude than the semipartial correlation between tenure profile fit scores and tenure partialing out the improbable item lie scale scores from the tenure profile fit scores ($r = -.18$). This difference ($r_d = .02$) was small but statistically significant, $t(953) = -3.15, p < .05$. Since no difference was expected, Hypothesis 4a was not supported. When corrected on the basis of the consistency-in-

responding lie scale scores, the corrected correlation was $-.16$, which is identical to the uncorrected correlation (i.e., $r = -.16$). This difference was not statistically significant, $t(953) = 0.00, p > .05$. Thus, Hypothesis 4b was supported.

Hypothesis 5a stated that the difference between the uncorrected and corrected criterion-related validity of the linear composite scores in predicting tenure would be statistically significant. The relationship between the linear composite score and tenure was small ($r = .07$). When corrected on the basis of improbable item lie scale scores, the corrected linear composite score displayed a smaller relationship with tenure ($r = .06$) compared to the uncorrected linear composite scores. However, this difference ($r_d = .01$) was not statistically significant, $t(953) = 0.39, p > .05$. Thus, Hypothesis 5a was not supported. When corrected on the basis of consistency-in-responding lie scale scores, the criterion-related validity of the corrected linear composite score was $.07$, which is identical to the uncorrected criterion-related validity coefficient (i.e., $r = .07$). Thus, Hypothesis 5b was not supported.

Hypothesis 6a stated that the differences in corrected and uncorrected criterion-related validity will be larger for the linear composite score compared to the tenure profile fit score when these scores are corrected on the basis of the improbable item lie scale scores. The difference between the corrected and uncorrected criterion-related validity for the linear composite was minimal ($r_d = -.01$) and smaller than the difference between the corrected and uncorrected criterion-related validity of the tenure profile fit score ($r_d = -.02$). Thus, Hypothesis 6a was not supported.

The difference between the uncorrected criterion-related validity of the tenure profile fit score ($r = -.16$; see Table 5) and the criterion-related validity corrected for consistency-in-responding lie scale scores ($r = -.16$) was 0.00, where the difference between the uncorrected criterion-related validity of the linear composite score ($r = .07$) was identical to the corrected criterion-related validity of the linear composite score when the linear composite score was corrected on the basis of the consistency-in-responding lie scale score (i.e., $d_r = 0.00$). Correcting for the consistency-in-responding lie scale scores did not affect the criterion-related validity of either the tenure profile fit score or the linear composite scores. Thus, Hypothesis 6b was not supported.

Subgroup Differences. Table 6 presents the means, standard deviations and effect sizes for the profile fit and linear composite scores by race and sex. The race-based subgroup differences were computed using Whites as the comparator, such that a positive d indicates Whites scored higher than the other subgroup. For the sex-based subgroup differences, males were the comparator, such that a positive d indicates males scored higher than females. None of the comparisons were statistically significant. The pattern of results was similar across the voluntary turnover fit scores, tenure profile fit scores, and the linear composite scores.

Table 5

Means, Standard Deviations, and Correlations for Study Variables

Variables	Mean	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Voluntary Turnover Profile Fit	5.31	4.86															
2. Tenure Profile Fit	5.27	4.59	0.97														
3. Simulated Profile 1 Fit	5.06	4.86	0.97	0.96													
4. Simulated Profile 2 Fit	5.44	4.91	0.99	0.96	0.98												
5. Simulated Profile 3 Fit	6.57	5.10	0.97	0.93	0.94	0.99											
6. Linear Composite	-0.02	3.72	0.21	0.09	0.35	0.31	0.25										
7. Agreeableness	54.42	16.40	0.04	0.02	0.23	0.18	0.13	0.79									
8. Conscientiousness	54.14	14.88	0.22	0.14	0.35	0.30	0.23	0.88	0.67								
9. Emotional Stability	55.57	14.90	0.06	0.03	0.26	0.13	0.00	0.82	0.74	0.71							
10. Extraversion	52.09	10.39	0.12	0.02	0.25	0.20	0.16	0.79	0.48	0.63	0.57						
11. Openness	43.78	8.68	0.35	0.12	0.22	0.32	0.41	0.42	0.03	0.23	0.02	0.26					
12. Consistency-in-Responding	0.11	0.11	0.05	0.05	0.01	0.03	0.06	-0.09	-0.15	0.09	-0.17	-0.11	0.03				
13. Improbable Item	2.68	2.61	0.27	0.22	0.40	0.34	0.27	0.73	0.55	0.76	0.62	0.64	0.15	0.11			
14. Tenure (Days)	190.28	167.62	-0.15	-0.16	-0.12	-0.15	-0.16	0.07	0.06	-0.02	0.13	0.15	-0.05	-0.09	0.05		
15. Involuntary Turnover	0.08	0.27	-0.06	-0.06	-0.04	-0.06	-0.07	0.05	0.04	0.02	0.07	0.08	-0.03	-0.05	0.02	-0.04	
16. Voluntary Turnover	0.31	0.46	-0.06	-0.06	-0.04	-0.05	-0.05	0.07	0.08	0.05	0.06	0.08	-0.03	-0.04	0.05	-0.30	-0.20

Note. $N = 956$; $r > .05$ are statistically significant at $p < .05$ (one-tailed).

Profile Scatter

Susceptibility to Response Distortion. A secondary objective of the current study was to examine the effects of profile configuration (or scatter) on the efficacy of profile

Table 6

Subgroup Differences for Profile Fit and Linear Composite Scores

Subgroup	Mean	SD	N	d
Voluntary Turnover Profile Fit				
African–American	5.34	4.61	204	-0.04
American Indian	8.32	6.66	7	-0.56
Asian	6.81	6.64	18	-0.29
Hawaiian/Pacific Islander	3.32	1.68	6	0.54
Hispanic	5.37	5.12	476	-0.05
White	5.15	4.45	211	—
Female	5.09	4.77	543	0.11
Male	5.64	4.99	408	—
Tenure Profile Fit				
African–American	5.48	4.57	204	-0.12
American Indian	8.13	7.11	7	-0.55
Asian	5.88	5.68	18	-0.19
Hawaiian/Pacific Islander	3.32	1.68	6	0.52
Hispanic	5.34	4.77	476	-0.09
White	4.95	4.12	211	—
Female	5.13	4.59	543	0.08
Male	5.48	4.61	408	—
Linear				
African–American	-0.29	3.77	204	0.14
American Indian	-2.32	3.90	7	0.69
Asian	2.33	3.84	18	-0.56
Hawaiian/Pacific Islander	-1.92	2.09	6	0.71
Hispanic	-0.04	3.78	476	0.08
White	0.24	3.56	211	—
Female	-0.23	3.63	543	0.12
Male	0.23	3.85	408	—

Note. * $p < .05$ (two-tailed).

matching in reducing the effects of response distortion. Table 5 presents the means, standard deviations, and correlations for the simulated standard profiles. Simulated Profile 1 had the least amount of scatter and its profile fit scores displayed a statistically significant correlation with the improbable item lie scale score ($r = .40, p < .05$), such that test-takers who distorted their responses had worse fit compared to those who responded honestly. Simulated Profile 2 had the second least amount of scatter and its profile fit scores also significantly correlated with the improbable item lie scale scores ($r = .34, p < .05$). Hypothesis 7a, which stated that a standard profile with no scatter would display larger correlations with the improbable item lie scale scores compared to a standard profile with moderate scatter, was supported as the difference between these two correlations was statistically significant, $t(953) = 10.54, p < .05$. Simulated Profile 1 did not display a significant correlation with the consistency-in-responding lie scale scores ($r = .01, p > .05$). Simulated Profile 2 also did not significantly correlate with the consistency-in-responding lie scale scores ($r = .03, p < .05$). Thus, Hypothesis 7b, which stated the standard profile with no scatter would display larger correlations with the consistency-in-responding lie scale scores compared to the standard profile with moderate scatter, was not supported.

Simulated Profile 3 had the most amount of scatter of the three simulated profiles, and its profile fit scores displayed a moderate correlation with improbable item lie scale scores ($r = .27, p < .05$). Hypothesis 8a, which stated that profile fit scores based on a standard profile with a moderate amount of scatter would display larger correlations with the improbable item lie scale scores than profile fit scores based on a

standard profile with high scatter, was supported, $t(953) = 18.79, p < .05$. Simulated Profile 3 fit scores were not related to the consistency-in-responding lie scale scores ($r = .06, p < .05$). Hypothesis 8b, which stated that profile fit scores based on a standard profile with a moderate amount of scatter would display larger correlations with the consistency-in-responding lie scale scores than profile fit scores based on a standard profile with high scatter, was not supported.

Criterion-Related Validity. Hypothesis 9a stated that a standard profile with high scatter would display a larger criterion-related validity coefficient compared to a profile with a moderate amount of scatter. Simulated Profile 3 displayed a statistically significant correlation with tenure ($r = -.16$) which was larger than that displayed by Simulated Profile 2 ($r = -.15$) and the difference was statistically significant, $t(953) = -2.22, p < .05$. Thus, Hypothesis 9a was supported. Simulated Profile 1 also displayed a significant relationship with tenure ($r = -.12$) which was significantly smaller than that of Profile 2 and tenure, $t(953) = -4.73, p < .05$. Thus, Hypothesis 9b, which stated that a standard profile with moderate scatter would display a larger criterion-related validity coefficient compared to a profile with no scatter, was supported.

Hypothesis 10a stated that a standard profile with high scatter would display a minimal difference between corrected and uncorrected criterion-related validity coefficients in predicting tenure when the corrections were based on the improbable item lie scale scores. The uncorrected criterion-related validity coefficient was $-.16$, whereas the corrected validity coefficient was $-.18$. The difference between these two correlations was statistically significant, $t(953) = 2.22, p > .05$. Thus, Hypothesis 10a

was not supported. When corrected on the basis of the consistency-in-responding lie scale scores, the criterion-related validity was -.16. Thus, Hypothesis 10b, which stated the standard profile with high scatter would display minimal differences between corrected and uncorrected criterion-related validity coefficients in predicting tenure when corrected on the basis of the consistency-in-responding lie scale scores, was supported.

Hypothesis 11a stated that a standard profile with moderate scatter would display a practically significant difference between corrected and uncorrected criterion-related validity coefficients in predicting tenure when the corrections were based on the improbable item lie scale scores. The uncorrected criterion-related validity coefficient was -.15, whereas the corrected validity coefficient was -.17. The difference between these two correlations was statistically significant, $t(953) = 1.81, p < .05$. Thus, Hypothesis 11a was supported. When corrected on the basis of the consistency-in-responding lie scale scores, the criterion-related validity was -.14, $t(953) = -2.21, p < .05$. Thus, Hypothesis 11b, which stated the standard profile with moderate scatter would display significant differences between corrected and uncorrected criterion-related validity coefficients in predicting tenure when corrected on the basis of the consistency-in-responding lie scale scores, was supported.

Hypothesis 12a stated that a standard profile with no scatter would display a meaningful difference between corrected and uncorrected criterion-related validity coefficients in predicting tenure when the corrections were based on the improbable item lie scale scores. The uncorrected criterion-related validity coefficient was -.12, whereas

the corrected validity coefficient was $-.15$. The difference between these two correlations was statistically significant, $t(953) = 2.21, p < .05$. Thus, Hypothesis 12a was supported. When corrected on the basis of the consistency-in-responding lie scale scores, the criterion-related validity was $-.12, t(953) = -0.00, p > .05$. Thus, Hypothesis 12b, which stated the standard profile with no scatter would display meaningful differences between corrected and uncorrected criterion-related validity coefficients in predicting tenure when corrected on the basis of the consistency-in-responding lie scale scores, was not supported.

Hypothesis 13a stated that the difference between the uncorrected and corrected validities would be smaller for the profile fit scores with high levels of scatter compared to the standard profile fit scores with moderate scatter when the corrections are based on the improbable item lie scale scores. The difference between the uncorrected and corrected criterion-related validities was $.02$ and $.02$ for the high and moderate scatter profile fit scores, respectively. Thus, Hypothesis 13a was not supported. When corrected on the basis of the consistency-in-responding lie scale scores, the difference between the uncorrected and corrected criterion-related validities was $.00$ and $-.01$. A 95% confidence interval for a correlation of $.01$ is $-.07$ to $.05$. Thus, Hypothesis 13b was not supported.

Hypothesis 14a stated that the difference between the uncorrected and corrected validities would be smaller for the profile fit scores with moderate scatter compared to the standard profile fit scores with no scatter when the corrections are based on the improbable item lie scale scores. The difference between the uncorrected and corrected

criterion-related validities was .02 and .03 for the moderate and no scatter profile fit scores, respectively. A 95% confidence interval for a correlation of .02 is -.04 to .08. Thus, Hypothesis 14a was not supported. When corrected on the basis of the consistency-in-responding lie scale scores, the difference between the uncorrected and corrected criterion-related validities was -.01 and .00. A 95% confidence interval for a correlation of .01 is -.07 to .05. Thus, Hypothesis 14b was not supported. Table 7 presents a summary of the research hypotheses and results.

Table 7

Summary of Research Hypotheses and Results

Hypothesis	Result
1. The observed correlation between the tenure profile fit score and linear composite will be modest ($r < .50$).	Supported
2a. The observed correlation between tenure profile fit scores and the improbable item lie scale scores will be smaller than that of the linear composite scores and the improbable item lie scale scores.	Supported
2b. The observed correlation between tenure profile fit scores and the consistency-in-responding lie scale scores will be smaller than that of the linear composite scores and the consistency-in-responding lie scale scores.	Cannot be meaningfully interpreted
3a. The criterion-related validity of the tenure profile fit scores in predicting tenure will be larger than that of the linear composite scores.	Supported
3b. The criterion-related validity of the voluntary turnover profile fit scores in predicting voluntary turnover will be larger than that of the linear composite scores.	Not Supported
4a. The differences in corrected and uncorrected criterion-related validity of the tenure profile fit scores in predicting tenure will be minimal when profile fit scores are corrected on the basis of the improbable item lie scale scores.	Not Supported
4b. The differences in corrected and uncorrected criterion-related validity of the tenure profile fit scores in predicting tenure will be minimal when profile fit scores are corrected on the basis of the consistency-in-responding lie scale scores.	Supported
5a. The differences in corrected and uncorrected criterion-related validity of the linear composite scores in predicting tenure will be significant when linear composite scores are corrected on the basis of the improbable item lie scale scores.	Not Supported
5b. The differences in corrected and uncorrected criterion-related validity of the linear composite scores in predicting tenure will be significant when linear composite scores are corrected on the basis of the consistency-in-responding lie scale scores.	Not Supported

Table 7 (cont).

Hypothesis	Result
6a. The differences in corrected and uncorrected criterion-related validity will be larger for the linear composite score compared to the profile fit score when these scores are corrected on the basis of the improbable item lie scale scores.	Not Supported
6b. The differences in corrected and uncorrected criterion-related validity will be larger for the linear composite score compared to the profile fit score when these scores are corrected on the basis of the consistency-in-responding lie scale scores.	Not Supported
7a. The profile fit scores based on an ideal profile with no scatter will display a larger correlation with the improbable item lie scale scores compared to the profile fit scores based on a profile with moderate levels of scatter.	Supported
7b. The profile fit scores based on an ideal profile with no scatter will display a larger correlation with the consistency-in-responding lie scale scores compared to the profile fit scores based on a profile with moderate levels of scatter.	Not Supported
8a. The profile fit scores based on an ideal profile with moderate levels of scatter will display larger correlations with the improbable item lie scale scores compared to the profile fit scores based on a profile with high levels of scatter.	Supported
8b. The profile fit scores based on an ideal profile with moderate levels of scatter will display larger correlations with the consistency-in-responding lie scale scores compared to the profile fit scores based on a profile with high levels of scatter.	Not Supported
9a. The profile fit scores based on a standard profile with high levels of scatter will display a larger criterion-related validity coefficient compared to the profile fit scores based on a standard profile with moderate levels of scatter.	Supported
9b. The profile fit scores based on a standard profile with moderate scatter will display a larger criterion-related validity coefficient compared to the profile fit scores based on a standard profile with no scatter.	Supported

Table 7 (cont).

Hypothesis	Result
10a. The profile fit scores based on a standard profile with high levels of scatter will display a minimal difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the improbable item lie scale.	Not Supported
10b. The profile fit scores based on a standard profile with high levels of scatter will display a minimal difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the consistency-in-responding lie scale scores.	Supported
11a. The profile fit scores based on a standard profile with moderate levels of scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the improbable item lie scale.	Supported
11b. The profile fit scores based on a standard profile with moderate levels of scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the consistency-in-responding lie scale scores.	Supported
12a. The profile fit scores based on a standard profile with no scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the improbable item lie scale.	Supported
12b. The profile fit scores based on a standard profile with no scatter will display a practically significant difference between corrected and uncorrected criterion-related validities in predicting tenure when corrections are based on the consistency-in-responding lie scale scores.	Not Supported
13a. The differences between corrected and uncorrected criterion-related validities will be smaller for profile fit scores based on the standard profile with high levels of scatter compared to that of the profile fit scores based on a standard profile with moderate levels of scatter when corrections are based on the improbable item lie scale.	Not Supported

Table 7 (cont).

Hypothesis	Result
13b. The differences between corrected and uncorrected criterion-related validities will be smaller for profile fit scores based on the standard profile with high levels of scatter compared to that of the profile fit scores based on a standard profile with moderate levels of scatter when corrections are based on the consistency-in-responding lie scale scores.	Not Supported
14a. The difference between corrected and uncorrected criterion-related validities will be smaller for the profile fit scores based on a standard profile with moderate levels of scatter compared to that of the profile fit scores based on a standard profile with no scatter when corrections are based on the improbable item lie scale.	Not Supported
14b. The difference between corrected and uncorrected criterion-related validities will be smaller for the profile fit scores based on a standard profile with moderate levels of scatter compared to that of the profile fit scores based on a standard profile with no scatter when corrections are based on the consistency-in-responding lie scale scores.	Not Supported

CHAPTER VII

DISCUSSION AND CONCLUSION

The primary objective of the current study was to provide an initial proof-of-concept test of the efficacy of profile matching in reducing the effects of response distortion. Specifically, the current study compared profile fit scores and linear composite scores in terms of their relationship with response distortion, criterion-related validity in predicting tenure and voluntary turnover, and the extent to which correcting for response distortion influenced said criterion-related validities. The current study also sought to investigate the role of profile configuration in these effects, and document the extent of subgroup differences in personality profile fit scores.

A number of summary statements can be made on the basis of the results presented here. First, personality profile fit scores appear to represent a different manifestation of the personality domain space than the linear composite scores. Specifically, the correlations between the empirical profile fit scores (i.e., tenure and voluntary turnover profile fit scores) and the linear composite scores were small. Furthermore, the relationships between the simulated profiles were slightly larger, but still small to moderate in magnitude. These results are commensurate with calls for person-centered or whole person approaches in investigating organizational behavior (Foti, Thompson, & Allgood, 2011; Weiss & Rupp, 2011a, 2011b).

Second, profile fit scores were less related to response distortion, as measured by the improbable item lie scales, compared to linear composite scores. The consistency–

in-responding lie scale demonstrated a small relationship with the tenure profile fit scores and this relationship was similar in magnitude to that of the linear composite score. Thus, the extent to which profile fit scores are susceptible to response distortion would seem to be dependent upon the type or operationalization of response distortion.

The difference in the pattern of relationships between the improbable item and consistency-in-responding lie scales could be due to the differences in the types of response distortion they detect. Specifically, the improbable item lie scale operationalizes response distortion as the tendency to endorse socially desirable items which are unlikely to be true. It is reasonable to posit that individuals who endorse a large number of improbable items also inflate their scores on the focal personality measures by inflating their responses via extreme responding. This would result in higher linear composite scores, but not necessarily higher profile fit scores. In order to inflate their profile fit scores, test-takers would have to know or guess the ideal profile being used and then respond to match it.

The conceptual relationships between the consistency-in-responding lie scale scores, profile fit scores, and linear composite scores are less clear. The consistency-in-responding lie scale operationalizes response distortion as the tendency to respond to similar items differently depending on their social desirability. Thus, unlike the improbable item lie scale, test-takers who distort their responses in this manner would be expected to receive lower scores on the linear composite after the items have been reverse-coded. It is important to note that there is some controversy in the extant literature as to the meaning of consistency-in-responding lie scales and it would seem

that they are not as widely used as improbable item lie scales to operationalize response distortion.

The next set of hypotheses examined the extent to which linear composite scores and profile fit scores displayed different criterion-related validities with tenure and voluntary turnover. The tenure profile fit score displayed a larger criterion-related validity with tenure compared to that of the linear composite score. However, the voluntary turnover profile fit score did not demonstrate larger correlations with voluntary turnover compared to the linear composite scores. Thus, it would seem that the benefit in using profile matching, in terms of criterion-related validity, depends upon the outcome of interest.

It is important to note that the test-takers in the current dataset were applicants who were hired in part on the basis of their personality test scores. The hiring organization used a different ideal profile and a proprietary method of assessing profile fit. As such, the observed effects may not represent the population correlation due to range restriction (see Sackett & Yang, 2000 for a review). The variance of both the profile fit scores used in the current study and the linear composite was probably reduced due to selection based on the proprietary profile. Thus, the observed correlations were lower than what would be expected if no range restriction occurred.

The tenure profile fit scores did demonstrate a higher criterion-related validity coefficient compared to the linear composite in predicting tenure. This difference could be due to the manner in which personality was operationalized, or to their differential susceptibility to response distortion. When corrected for response distortion, as

measured by the improbable item lie scale, the criterion-related validity of the tenure profile fit scores was significantly larger, suggesting that response distortion did attenuate the relationship between profile fit scores and tenure. However, correcting the linear composite score for response distortion, as measured by the improbable item lie scale, did not result in higher criterion-related validity. This suggests that response distortion did not attenuate the relationship between the linear composite scores and tenure.

When response distortion was measured by the consistency-in-responding lie scale scores, the corrected and uncorrected criterion-related validities were identical for both the tenure profile fit scores and the linear composite. Response distortion, as operationalized by the consistency-in-responding lie scale scores, did not appear to attenuate the criterion-related validity of the profile fit or linear composite scores.

The zero-order correlations between the improbable item lie scale and tenure profile fit scores and the results obtained when the tenure profile fit scores were corrected on the basis of the improbable item lie scale scores seem to be at odds with one another. Specifically, the zero-order correlations suggest that profile fit scores are less susceptible to response distortion, as measured by the improbable item lie scale, compared to the linear composite score. On the other hand, correcting for response distortion had a larger effect on the profile fit scores compared to the linear composite score. One explanation is that correcting for response distortion is not an effective method for reducing the effects of response distortion. Several researchers have investigated the efficacy of statistical controls for response distortion (e.g., Hough, 1998;

Ones et al., 1996). Hough (1998) found that correcting scale scores based on the Unlikely Virtues scale did not affect criterion-related validity. However, said corrections resulted in mean scores that were more commensurate with incumbent (as opposed to applicant) mean scores, and resulted in different hiring recommendations (see also Rosse et al., 1998). Furthermore, Zickar (2000) argued that the correlation coefficient may not be an effective metric for detecting the effect of response distortion as the correlation coefficient is not sensitive to disruptions in the bivariate distribution.

The current study also sought to investigate the role of profile configuration in the efficacy of profile matching in reducing the effects of response distortion. It was posited that standard profiles with lower levels of scatter would display stronger relationships with response distortion compared to standard profiles with higher levels of scatter. Three simulated profiles were used—profiles with high, moderate, and no scatter. The results suggest that profile fit scores based on a standard profile with no scatter displayed a stronger relationship with response distortion, as measured by the improbable item lie scale, compared to profile fit scores based on a standard profile with moderate scatter. Furthermore, profile fit scores based on a standard profile with moderate scatter displayed a stronger relationship with response distortion compared to profile fit scores based on a standard profile with high levels of scatter. Thus, when the improbable item lie scale was used to operationalize response distortion, it appears that profile configuration was negatively associated with response distortion. It is important to note that these relationships did not hold when consistency-in-responding lie scales

were used. This may be due to the construct-related validity concerns regarding the use of consistency-in-responding scores to operationalize response distortion.

The level of scatter was also positively related to the criterion-related validity of the profile fit scores. Specifically, profile fit scores based on a profile with more scatter displayed stronger correlations with tenure than that of fit scores based on profile fit scores with moderate levels of scatter. Furthermore, profile fit scores based on a standard profile with moderate scatter displayed stronger relationships with turnover compared to that of profiles with no scatter.

The differences in criterion-related validity did not appear to be attributable to response distortion. Specifically, correcting for response distortion had a significant effect on the criterion-related validities of all three simulated profiles, when response distortion was measured by the improbable item lie scale. When response distortion was measured by the consistency-in-responding lie scale, the pattern of results was difficult to interpret. Specifically, the profiles with no and large amounts of scatter did not appear to be susceptible to response distortion. However, the profile with a moderate level of scatter was susceptible to response distortion.

In addition to investigating the efficacy of profile matching in reducing the effects of response distortion, the current study sought to document the extent to which personality profile fit scores displayed subgroup differences, if any. None of the subgroup differences were statistically significant. It should be noted that some comparisons resulted in large effect sizes, but because these effect sizes were associated with rather small sample sizes, they were not statistically significant. Although not

statistically significant, the pattern of results was quite similar across the two empirical profile fit scores (i.e., tenure and voluntary turnover profile fit). However, there were some notable differences between the profile fit scores and the linear composite scores. Specifically, the linear composite scores indicated that all subgroups, with the exception of Asians, scored higher than Whites. Conversely, for the profile fit scores, all subgroups, with the exception of Asians, scored lower (i.e., better fit) than Whites. As a general summary statement, it appears that the initial evidence suggests that profile fit scores do not introduce subgroup difference concerns into personality test scores.

Critiques of Profile Matching

Although profile matching has not received much, if any, empirical attention in the published I/O personality literature, it is widely used by several testing firms. The use of profile matching is controversial, with most of the critiques being conceptual rather than empirical. That is, although there is a dearth of empirical evidence comparing profile matching to linear approaches, there is a debate as to the appropriateness of this method for employment-related decision making. The critiques from one firm will be reviewed here (SHL PreVisor, 2011), and counter points to their critiques will also be discussed.

Profile matching approaches have been critiqued on the basis of the use of small samples for the referent group because the use of small referent groups threatens the generalizability of the profile. However, the use of very small samples (e.g., six incumbents) threatens the veracity of other approaches as well (e.g., content and criterion validation designs). Thus, like other approaches, the available sample should be

considered when deciding whether a profile matching strategy is appropriate. Similarly, profile matching has been critiqued for not using job analyses to determine job–relevant and job–irrelevant traits. However, the process by which standard profiles are developed by definition, inherently identify job–relevant traits. Specifically, only variables that distinguish high performing employees from low performing employees are included in the profile. Furthermore, there is no reason why a traditional job analysis could not be used in conjunction with profile matching and the use of a traditional job analysis would reduce the likelihood of using proxy variables that are job–irrelevant but covary with job–relevant variables.

The other critiques outlined in the SHL PreVisor’s (2011) white paper are more germane to a discussion of the efficacy of profile matching. A particularly insightful critique is the so–called ceiling effect problem. This problem asserts that profile matching makes it impossible to identify and select test–takers who are a better performers than the employees contained in the referent sample. That is, profile matching engenders creating a profile which represents an ideal employee and assessing applicants in terms of how well their profile matches the ideal profile. Since the ideal profile is developed using current employees, the use of profile matching will never result in better employees than those already in the organization.

A related critique is that profile matching results in unnecessarily high homogeneity amongst employees. The attraction–selection–attrition (ASA) cycle proposes that three processes—attraction, selection, and attrition—result in organizations that are homogenous in terms of employees’ personalities (Schneider,

1987). Profile matching would seem to accelerate this process. Organizational homogeneity may result in groupthink (Janis, 1982) and lead to organizations that are unable to adapt (Schneider, 1987). Furthermore, heterogeneous teams and groups have the potential to generate more high-quality decisions and consider a greater range of perspectives when problem-solving (Watson, Kumar & Michaelsen, 1993). Thus, the cloning problem could have detrimental effects for organizations that use profile matching. The cloning problem could be mitigated by only using job-relevant traits in the ideal profile. However, it is unlikely that the use of job-relevant traits would completely remove this potential problem.

A final critique that will be discussed here is the lower reliability of profile fit scores compared to raw scores. The use of difference scores results in lower reliabilities than the use of raw scores. However, the use of difference scores does not necessarily result in reliability that is below the standard cutoffs for the use of test in making personnel decisions.

Limitations and Directions for Future Research

One potential limitation of the current study is the personality measures used. Specifically, these measures display intercorrelations between the FFM dimensions that are higher than what is generally observed in the extant literature. The intercorrelations range from 0.02 to 0.74. Thus, some of the dimensions display correlations that are more commensurate with alternative forms reliability than the typical intercorrelations between FFM dimensions. Table 8 presents the intercorrelations amongst the FFM variables in the present study and meta-analytic true score estimates reported by Ones

(1993). In all but two cases, the intercorrelations from the present study are much larger than that reported by Ones.

Table 8

Intercorrelations Amongst FFM Variables Reported by Ones (1993) and From the Current Study

Variable	1.	2.	3.	4.
1. Emotional Stability				
2. Extraversion	.19/.57			
3. Openness	.16/.02	.17/.26		
4. Agreeableness	.25/.74	.17/.48	.11/.03	
5. Conscientiousness	.26/.71	.00/.63	-.06/.23	.27/.67

Note. The first correlation in each cell are meta-analytic true score estimates as reported by Ones (1993) and the second correlation in each cell is the zero-order correlations from the current study.

This is particularly problematic as the effectiveness of using test score profiles is inversely related to the intercorrelations between the variables contained in the profile. This is evident in the high correlations between the profile fit scores. For example, one would expect a fairly low correlation between the tenure profile fit scores and the voluntary turnover fit scores. The former was based on employees who remained with the organization for more than one year, whereas the latter was based on employees who voluntarily turned over within one month. Although one would expect these profile fit scores to display a small, or even negative, correlation, the relationship was quite strong ($r = .97$). An examination of the empirical standard profiles suggests that these two profiles are not largely differentiated—they correlate .99 (Figure 8 presents the standard

profiles for voluntary turnover and tenure). The resultant critical question is whether these two standard profiles are meaningfully distinguishable. Although these two profiles appear similar, they are not identical. Specifically, the tenure profile had higher scale scores for all variables with the exception of openness. This is commensurate with Zimmerman (2008) who found that the FFM variables were negatively correlated with turnover, with the exception of openness which was positively related to turnover.

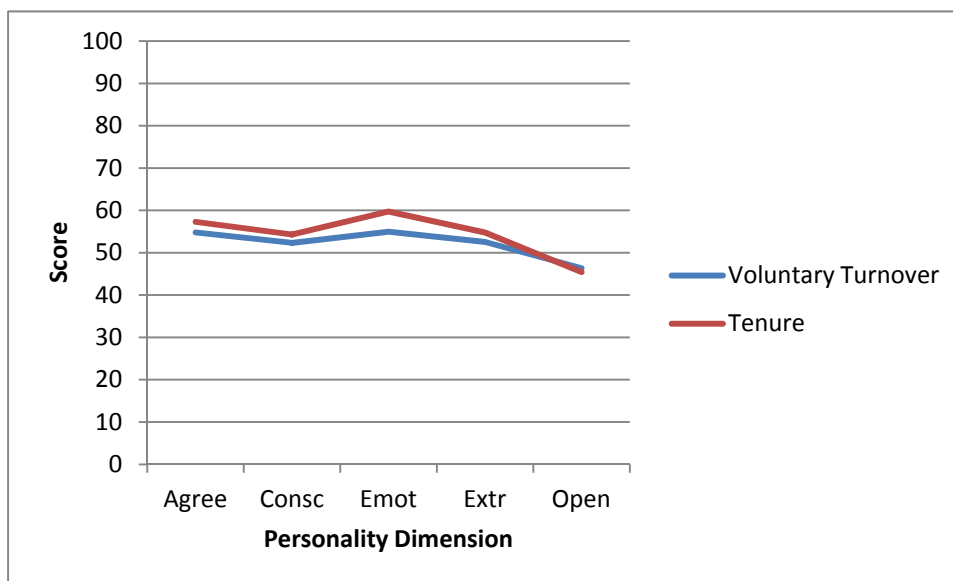


Figure 8. Standard profiles for voluntary turnover and tenure.

The current study used two operationalizations of response distortion—an improbable item lie scale and a consistency-in-responding lie scale. The former is often used in applied settings (Goffin & Christiansen, 2003) ostensibly due to its conceptual foundation. That is, these items are designed to be highly socially desirable but unlikely to be true. However, these types of scales are vulnerable to false positives (McCrae &

Costa, 1983) and false negatives (Dwight & Alliger, 1997). Consistency-in-responding lie scales appear less common and their construct-related validity is questionable (cf. Bond, 1986; Clark et al., 2003). The results of the current study should be interpreted within the limitations that these methods of detecting response distortion present. The results where the improbable item lie scale was used to operationalize response distortion display consistent results—profile fit scores are less related to response distortion, but their criterion-related validities are affected by correcting for response distortion. The results are less clear when the consistency-in-responding approach is used. The general pattern of results is more favorable for the improbable item lie scale than for the consistency-in-responding lie scale and this is consonant with their use in the extant literature. Future research should use different operationalizations of response distortion, including idiosyncratic faking approaches (Kuncel & Borneman, 2007), response latencies (Zuckerman et al., 1981), eye-tracking (Van Hoff & Born, 2012), and other innovative approaches to detecting response distortion.

The present study used data that were collected in an unproctored internet-based manner. Although unproctored internet-based testing has become commonplace, researchers and practitioners are concerned with the validity of test scores collected using this method, especially in high-stakes testing. One threat that is particularly germane to the present study is the use of surrogate test-takers. If surrogates were used, then the personality test scores would reflect the responses of one individual, whereas the criterion scores would reflect the behavior of another individual. However, the presence of a proctor is not a means of deterring or detecting response distortion on

noncognitive measures (Arthur & Glaze, 2011). Thus, it is unlikely that the results presented here are particular to unproctored testing.

The current study used tenure/turnover as the criterion and future studies should explore other organizationally relevant outcomes. The relationship between personality variables and turnover is quite small in the extant literature (Zimmerman, 2008). Turnover can result from a myriad of antecedents such as the work environment (Griffeth et al., 2000), attitudes (Johns, 2002), as well as personality (Zimmerman, 2008). Given the number of possible causes and interactions, it is unlikely that any one variable or class of variables will display large criterion-related validities. Furthermore, predicting turnover is difficult given the dichotomous nature of the data and low base rates. Finally, the conceptual link between personality and turnover is less clear than the relationship between personality and job performance. It is likely, given the different relationships between personality and job performance, and response distortion and job performance, that the efficacy of profile matching in reducing the effects of response distortion would be greater when job performance is used as a criterion. Thus, future research should investigate the efficacy of profile matching approaches in predicting various other criteria.

Finally, the present study consisted of a blend of real and simulated data. Thus, the study engenders some of the advantages and disadvantages of both a field test and a simulation. Simulations are fairly common in I/O psychology (see Roth, Switzer, Van Iddekinge, & Oh, 2011) especially in the adverse impact literature (e.g., Biddle, & Morris, 2011; Doverspike, Winter, Healy, & Barrett, 1996; Dunlevy, Mueller,

Buonasera, Kuang, & Dunleavy, 2008; Finch, Edwards, & Wallace, 2009; Tam, Murphy, & Lyall, 2004) and the faking and response distortion literature (Berry & Sackett, 2009; Converse, Peterson, & Griffith, 2009; Komar, Brown, Komar, & Robie, 2008; Schmitt & Oswald, 2006). Simulations have a number of advantages, including their (a) ability to allow for investigation of phenomena that are difficult or impossible to observe in field or laboratory tests, (b) precision and control, (c) theoretical rigor, and (d) cost-effectiveness. However, simulations engender a number of limitations and disadvantages, including a lack of generalizability and difficulty in execution. Specifically, simulations are based on statistical inputs (usually from primary or meta-analytic studies) and not from observed behavior. Furthermore, simulations are reductionist, in that they may exclude relevant variables (Zickar & Slaughter, 2002). Field tests allow for hypotheses to be tested under realistic conditions. Given the extent of missing data, the relationship between the focal personality variables and the two lie scales was estimated based on a subset of the data and the missing data were then imputed. The use of operational field data to estimate the relationship between the focal personality variables and the lie scales provides a realistic estimate of these relationships. However, because a mix of real and simulated data were used, it is particularly important that the present study be replicated in a field test.

Practical and Scientific Implications

The primary impetus for the current study was to empirically evaluate profile matching as a personnel decision-making strategy. Although the extant literature in this domain is quite limited, this approach is extensively used by testing and assessment

firms. Thus, the present study addressed the appropriateness and effectiveness of using profile matching to make personnel decisions especially within the context of alleviating concerns about response distortion. The results of the current study would seem to provide initial evidence of the efficacy of a profile matching approach. Profile fit scores displayed smaller correlations with response distortion (as measured by the improbable item lie scale) and larger criterion-related validities in predicting tenure. Furthermore, the results of the present study suggest that profile configuration plays a role in both the weaker association between profile fit scores and response distortion and the stronger criterion-related validities of profile fit scores. Thus, when response distortion and criterion-related validity are of concern, profile matching may be preferable over personnel decision making strategies that are based on a linear model.

The present study contributes to the personality and response distortion literature in several ways. First, the results suggest that profile fit scores represent a different manifestation of personality scores than linear composites. Several studies have investigated the joint effects of multiple personality variables (e.g., Judge & Erez, 2007; Warr et al., 2005; Witt et al., 2002), but all of these studies examined only two personality variables. In contrast, the current study used all five FFM variables.

Similar to other studies that have examined the efficacy of correcting personality test scores for response distortion (e.g., Hough, 1998; Ones et al., 1996), the current study found little to no effect for said correction for both the profile fit and linear composite scores. These findings are consistent with the observation that the correlation coefficient is not sensitive to distortions of the bivariate distribution (Zickar, 2000).

Thus, the current study adds to the extant literature which suggests correcting for response distortion may not be effective in addressing concerns regarding response distortion.

REFERENCES

- Ackerman, P. L., & Humphreys, L. G. (1990). *Individual differences in industrial and organizational psychology*. Palo Alto, CA, US: Consulting Psychologists Press.
- Alliger, G. M. (1988). Do zero correlations really exist among measures of difference cognitive abilities? *Educational and Psychological Measurement*, 48, 275-280. doi: 10.1177/0013164488482001
- Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, 60, 59-72. doi:10.1177/00131640021970367
- Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science*, 7, 32-39. doi:10.1111/j.1467-9280.1996.tb00663.x
- Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, 69, 574-580. doi:10.1037/0021-9010.69.4.574
- Arthur, W. Jr., & Glaze, R. M. (2011). Cheating and response distortion on remotely delivered assessments. In N. T. Tippins & S. Adler (Eds.), *Technology-enhanced assessment of talent* (pp. 99-152). San Francisco, CA: Jossey-Bass.
- Arthur, W. Jr., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18, 1-16. doi:10.1111/j.1468-2389.2010.00476.x

- Arthur, W. Jr., Villado, A. J., & Glaze, R. M. (2007). *Evaluating the effectiveness and efficacy of the PPI Falsification Scale: Recommendations for its use in the PeopleAnswers Assessment System*. College Station, TX: Winfred Arthur, Jr. Consulting.
- Arthur, W. Jr., Woehr, D. J., & Graziano, W. G. (2001). Personality testing in employment settings: Problems and issues in the application of typical selection practices. *Personnel Review*, 30, 657-676. doi:10.1108/EUM0000000005978
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81, 261-272. doi:10.1037/0021-9010.81.3.261
- Barrick, M. R., Mount, M. K., & Strauss, J. P. (1994). Antecedents of involuntary turnover due to a reduction in force. *Personnel Psychology*, 47, 515-536. doi:10.1111/j.1744-6570.1994.tb01735.x
- Bass, B. M. (1956). Development and evaluation of a scale for measuring social acquiescence. *Journal of Abnormal Social Psychology*, 53, 296-299. doi:10.1037/h0047161
- Benson, M. J., & Campbell, J. P. (2007). To be, or not to be, linear: An expanded representation of personality and its relationship to leadership performance.

International Journal of Selection and Assessment, 15, 232-249.

doi:10.1111/j.1468-2389.2007.00384.x

Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology*, 62, 835-863. doi:10.1111/j.1744-6570.2009.01159.x

Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301. doi:10.1111/j.1744-6570.2007.00074.x

Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random response indices: Validation using a self-report methodology. *Psychological Assessment*, 4, 340-345.

Biddle, D. A., & Morris, S. B. (2011). Using lancaster's mid-P correction to the fisher's exact test for adverse impact analyses. *Journal of Applied Psychology*, 96, 956-965. doi:10.1037/a0024223

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317-335. doi: 10.1111/j.1468-2389.2006.00354.x

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187-215. doi: 10.1037/0033-2909.117.2.187

Block, J. (1965). *The challenge of response sets*. New York: Appelton-Century-Croft.

- Bond, J. A. (1987). The process of responding to personality items: Inconsistent responses to repeated presentation of identical items. *Personality and Individual Differences*, 8, 409-417. doi: 10.1016/0191-8869(87)90042-0
- Bond, J. A. (1986). Inconsistent responding to repeated MMPI items: Is its major cause really carelessness? *Journal of Personality Assessment*, 50, 50-64. doi: 10.1207/s15327752jpa5001_7
- Bott, J. P., O'Connell, M. S., Ramakrishnan, M., & Doverspike, D. (2007). Practical limitations in making decisions regarding the distribution of applicant personality test scores based on incumbent data. *Journal of Business and Psychology*, 22, 123-134. doi:10.1007/s10869-007-9053-x
- Buechley, R., & Ball, H. (1952). A new test of "validity" for the group MMPI. *Journal of Consulting Psychology*, 16, 299-301.
- Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 35-38. doi:10.1111/j.1754-9434.2008.01104.x
- Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258-299). San Francisco: Jossey-Bass.
- Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawnsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the five factor model. *Journal of Applied Psychology*, 81, 694-704. doi:10.1037/0021-9010.81.6.694

- Chaplin, W. F. (2007). *Moderator and mediator models in personality research: A basic introduction*. New York, NY, US: Guilford Press.
- Chaplin, W. F. (1997). *Personality, interactive relations, and applied psychology*. San Diego, CA, US: Academic Press. doi:10.1016/B978-012134645-4/50034-2
- Chernyshenko, O. S., Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R., & Tuttle, M. D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: Empirical comparisons with other formats. *Human Performance*, 22, 105-127. doi:10.1080/08959280902743303
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18, 267-307. doi:10.1207/s15327043hup1803_4
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47, 847-860. doi:10.1111/j.1744-6570.1994.tb01581.x
- Clark, M. E., Gironda, R. J., & Young, R. W. (2003). Detection of back random responding: Effectiveness of MMPI-2 and Personality Assessment Inventory validity indices. *Psychological Assessment*, 15, 223-234. doi: 10.1037/1040-3590.15.2.223
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312. doi:10.1037/0003-066X.45.12.1304

- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, *16*, 155-169. doi:10.1111/j.1468-2389.2008.00420.x
- Converse, P. D., Peterson, M. H., & Griffith, R. L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment*, *17*, 47-60. doi:10.1111/j.1468-2389.2009.00450.x
- Cooper, D., & Robertson, I.T. (1995). *The psychology of personnel Selection: A quality approach*. New York: Routledge.
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*, 653-665. doi:10.1016/0191-8869(92)90236-I
- Costa, P. T., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, *55*, 258-265. doi:10.1037/0022-3514.55.2.258
- Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology*, *75*, 297-300. doi:10.1037/0021-9010.75.3.297
- Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology*, *80*, 532-537. doi:10.1037/0021-9010.80.4.532

- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684. doi:10.1037/h0043943
- Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50, 456-473. doi: 10.1037/h0057173
- Crowne, D. P., & Marlowe, D. A. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354. doi: 10.1037/h0047358
- Cucina, J. M., & Vasilopoulos, N. L. (2005). Nonlinear personality-performance relationships and the spurious moderating effects of traitedness. *Journal of Personality*, 73, 227-259. doi:10.1111/j.1467-6494.2004.00309.x
- Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14, 142-155. doi:10.1111/j.1468-2389.2006.00340.x
- Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology*, 79, 643-658. doi:10.1037/0021-9010.79.5.643
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Volume I: Clinical interpretation*. Minneapolis: University of Minnesota Press.

- Davison, M. L., & Davenport, E. C. Jr. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods*, 7, 486-484. doi: 10.1037//1082-989X.7.4.468
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, 42, 25-36. doi:10.1111/j.1744-6570.1989.tb01549.x
- De Raad, B., Hendriks, A. A. J., & Hofstee, W. K. B. (1994). The big five: A tip of the iceberg of individual differences. In C. F. Halverson Jr., G. A. Kohnstamm & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 91-109). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- De Raad, B., & Hofstee, W. K. (1993). A circumplex approach to the five factor model: A facet structure of trait adjectives supplemented by trait verbs. *Personality and Individual Differences*, 15, 493-505. doi:10.1016/0191-8869(93)90332-W
- Delery, J. E., & Kacmar, K. M. (1998). The influence of applicant and interviewer characteristics on the use of impression management. *Journal of Applied Social Psychology*, 28, 1649-1669. doi:10.1111/j.1559-1816.1998.tb01339.x
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440. doi: 10.1146/annurev.ps.41.020190.002221
- Dilchert, S. (2007). Peaks and valleys: Predicting interests in leadership and managerial positions from personality profiles. *International Journal of Selection and Assessment*, 15, 317-334. doi:10.1111/j.1468-2389.2007.00391.x

- Doverspike, D., Winter, J. L., Healy, M. C., & Barrett, G. V. (1996). Simulations as a method of illustrating the impact of differential weights on personnel selection outcomes. *Human Performance, 9*, 259-273.
- Dunlap, W. P., & Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research, 29*, 115-126. doi:10.1207/s15327906mbr2901_4
- Dunleavy, E. M., Mueller, L. M., Buonasera, A. K., Kuang, D. C., & Dunleavy, D. G. (2008). On the consequences of frequent applicants in adverse impact analyses: A demonstration study. *International Journal of Selection and Assessment, 16*, 333-344. doi:10.1111/j.1468-2389.2008.00439.x
- Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology, 15*, 13-24. doi:10.1111/j.1744-6570.1962.tb01843.x
- Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. *Educational and Psychological Measurement, 57*, 937-948. doi: 10.1177/0013164497057006004
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1-23. doi:10.1207/S15327043HUP1601_1
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Edwards, J. R. (2001a). Ten difference score myths. *Organizational Research Methods, 4*, 265-287. doi: 10.1177/109442810143005

- Edwards, J. R. (2001b). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350-400). San Francisco: Jossey-Bass.
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, 58, 51-100. doi: 10.1006/obhd.1994.1029
- Edwards, J. R., & Cooper, C. L. (1990). The person–environment fit approach to stress: Recurring problems and suggested solutions. *Journal of Organizational Behavior*, 11, 293-307. doi: 10.2307/256782
- Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, 36, 1577-1613. doi: 10.2307/256822
- Edwards, A. L., & Walsh, J. A. (1963). Relationship between various psychometric properties of personality items. *Educational and Psychological Measurement*, 23, 227-238. doi: 10.1177/001316446302300201
- Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, 92, 386-395. doi:10.1037/0021-9010.92.2.386
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84, 155-166. doi:10.1037/0021-9010.84.2.155

- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, 86, 122-133. doi:10.1037/0021-9010.86.1.122
- Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, 87, 1200-1208. doi:10.1037/0021-9010.87.6.1200
- Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology*, 86, 1270-1279. doi:10.1037/0021-9010.86.6.1270
- Finch, D. M., Edwards, B. D., & Wallace, J. C. (2009). Multistage selection strategies: Simulating the effects on adverse impact and expected performance for various predictor combinations. *Journal of Applied Psychology*, 94, 318-340. doi:10.1037/a0013775
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology*, 61, 579-616. doi:10.1111/j.1744-6570.2008.00123.x
- Frei, R. L., & McDaniel, M. A. (1998). Validity of customer service measures in personnel selection: A review of criterion and construct evidence. *Human Performance*, 11, 1-27. doi:10.1207/s15327043hup1101_1
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 9, 46-55. doi:10.1007/BF02686767

- Glenn, D. M., Beckham, J. C., Sampson, W. S., Feldman, M. E., Hertzberg, M. A., & Moore, S. D. (2002). MMPI-2 profiles for Gulf and Vietnam combat veterans with chronic posttraumatic stress disorder. *Journal of Clinical Psychology, 58*, 371-381. doi: 10.1002/jclp.1149
- Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340-344. doi:10.1111/j.0965-075X.2003.00256.x
- Goldberg, L. R. (1963). A model of item ambiguity in personality assessment. *Educational and Psychological Measurement, 23*, 467-492. doi: 10.1177/001316446302300305
- Gough, H. G., & Bradley, P. (1996). *CPI Manual* (3rd Ed). Palo Alto, CA: Consulting Psychologists Press.
- Gorman, B. S. (1968). Social desirability factors and the Eysenck Personality Inventory. *Journal of Psychology: Interdisciplinary and Applied, 69*, 75-83.
- Graziano, W. G., & Eisenberg, N. (1997). *Agreeableness: A dimension of personality*. San Diego, CA, US: Academic Press. doi:10.1016/B978-012134645-4/50031-7
- Graziano, W. G., Jensen-Campbell, L. A., & Hair, E. C. (1996). Perceiving interpersonal conflict and reacting to it: The case for agreeableness. *Journal of Personality and Social Psychology, 70*, 820-835. doi:10.1037/0022-3514.70.4.820
- Greene, R. L. (2000). *The MMP-2: An interpretive manual* (2nd ed). Needham Heights, MA: Allyn & Bacon.

- Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Clinical Psychology, 34*, 407-410.
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*, 463-488. doi:10.1177/014920630002600305
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*, 341-355. doi:10.1108/00483480710731310
- Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 308-311. doi:10.1111/j.1754-9434.2008.00053.x
- Guilford, J. P., Zimmerman, W. S., & Guilford, J. P. (1976). *The Guilford-Zimmerman Temperament Survey Handbook: Twenty-five years of research and application*. San Diego, CA: EdITS.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135-164. doi:10.1111/j.1744-6570.1965.tb00273.x
- Hand, J. (1964). Measurement of response sets. *Psychological Reports, 14*, 907-913.
- Hathaway, S. R., & Meehl, P. E. (1951). *An atlas for the clinical use of the MMPI*. Oxford, England: U. Minnesota Press.

- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683. doi:10.1111/j.1744-6570.2004.00003.x
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373-385. doi:10.1037/0021-9010.92.2.373
- Hechniger, J. (2008, September 11). Schools cancel GMAT Scores. *The Wall Street Journal*, p. D6.
- Heggstad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91, 9-24. doi:10.1037/0021-9010.91.1.9
- Hitt, M. A., & Barr, S. H. (1989). Managerial selection decision models: Examination of configural cue processing. *Journal of Applied Psychology*, 74, 53-61. doi:10.1037//0021-9010.74.1.53
- Hochwarter, W. A., Witt, L. A., Treadway, D. C., & Ferris, G. R. (2006). The interaction of social skill and organizational support on job performance. *Journal of Applied Psychology*, 91, 482-489. doi:10.1037/0021-9010.91.2.482
- Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63, 146-163. doi:10.1037/0022-3514.63.1.146

- Hogan, R. (1998). Reinventing personality. *Journal of Social and Clinical Psychology*, 17, 1-10. doi: 10.1521/jscp.1998.17.1.1
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270-1285. doi:10.1037/0021-9010.92.5.1270
- Hogan, R. T., & Hogan, J. (1992). *Hogan Personality Inventory Manual*. Tulsa, OK: Hogan Assessment Systems.
- Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality*, 23, 180-191. doi:10.1016/0092-6566(89)90022-6
- Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology*, 37, 379-386. doi:10.1002/1097-4679(198104)37:2<379::AID-JCLP2270370227>3.0.CO;2-6
- Holland, J. L. (1985). *Making vocational choices*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30, 161-172. doi:10.1177/0146167203259930
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209-244. doi: 10.1207/s15327043hup1102&3_6

- Hough, L. M. (1992). The "big five" personality variables—construct confusion: Description versus prediction. *Human Performance*, 5, 139-155.
doi:10.1207/s15327043hup0501&2_8
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.
doi:10.1037/0021-9010.75.5.581
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 272-290. doi:10.1111/j.1754-9434.2008.00048.x
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194.
doi:10.1111/1468-2389.00171
- Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, 53, 278–295. doi:10.1207/s15327752jpa5302_6
- Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence*, 3, 105-120.

- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362. doi:10.1016/0001-8791(86)90013-8
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98. doi:10.1037/0033-2909.96.1.72
- Hurtz, G. M., & Alliger, G. M. (2002). Influence of coaching on integrity test performance and unlikely virtues scale scores. *Human Performance*, 15, 255-273. doi:10.1207/S15327043HUP1503_02
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869-879. doi: 10.1037//0021-9010.85.6.869
- Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity and corporate financial performance. *Academy of Management Journal*, 38, 635–672. doi: 10.2307/256741
- Inwald, R. E. (1992). *Inwald Personality Inventory Technical Manual* (Revised Version 2A). New York: Hilson Research.
- Jackson, D. N. (1999). *Personality Research Form Manual* (3rd ed.). Port Huron, MI: Sigma Assessment Systems.
- Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston: Houghton Mifflin.
- Jenkins, C. A. (2002). *Improving the accuracy of personality measurement through profile-level judgments*. Unpublished doctoral dissertation. Hattiesburg, MS: University of Southern Mississippi.

- Johns, G. (2002). The psychology of lateness, absenteeism, and turnover. In N. Anderson, D. Ones, H. K. Sinangil, & C. Viswesvaran, (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 2, pp. 232–252). Thousand Oaks, CA: Sage.
- Johnson, J. A., & Ostendorf, F. (1993). Clarification of the five-factor model with the abridged big five dimensional circumplex. *Journal of Personality and Social Psychology*, 65, 563-576. doi:10.1037/0022-3514.65.3.563
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765-780. doi:10.1037/0021-9010.87.4.765
- Judge, T. A., & Erez, A. (2007). Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *Personnel Psychology*, 60, 573-596. doi:10.1111/j.1744-6570.2007.00084.x
- Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology*, 56, 303-331. doi:10.1111/j.1744-6570.2003.tb00152.x
- Kanfer, R., Ackerman, P. L., Murtha, T., & Goff, M. (1995). *Personality and intelligence in industrial and organizational psychology*. New York, NY, US: Plenum Press.
- Kaufman, R. A., Hakmiller, K. L., & Porter, L. W. (1959). The effects of top and middle management sets on the Ghiselli Self-Description Inventory. *Journal of Applied Psychology*, 43, 149-153. doi:10.1037/h0048890

- Kelley, P. L., Jacobs, R. R., & Farr, J. L. (1994). Effects of multiple administrations of the MMPI for employee screening. *Personnel Psychology*, 47, 575-591.
doi:10.1111/j.1744-6570.1994.tb01738.x
- Kluger, A. N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology*, 46, 763-780.
doi:10.1111/j.1744-6570.1993.tb01568.x
- Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A monte carlo investigation. *Journal of Applied Psychology*, 93, 140-154. doi:10.1037/0021-9010.93.1.140
- Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: A systematic evaluation. *Journal of Gerontology*, 42, 56-59.
- Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology*, 43, 48-55.
doi:10.1037/h0076266
- Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, 15, 220-231.
doi:10.1111/j.1468-2389.2007.00383.x
- Kuncel, N. R., & Tellegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology*, 62, 201-228. doi: 10.1111/j.1744-6570.2009.01136.x

- Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, 76, 315-332. doi: 10.1207/S15327752JPA7602_12
- LaHuis, D. M., Martin, N. R., & Avis, J. M. (2005). Investigating nonlinear conscientiousness-job performance relations for clerical employees. *Human Performance*, 18, 199-212. doi:10.1207/s15327043hup1803_1
- Lamb, D. G., Berry, D. T. R., Wetter, M. W., & Baer, R. A. (1994). Effects of two types of information on malingering of closed head injury on the MMPI-2: An analog investigation. *Psychological Assessment*, 6, 8-13. doi:10.1037/1040-3590.6.1.8
- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96, 202-210. doi:10.1037/a0020375
- Le, H., Oh, I., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96, 113-133. doi:10.1037/a0021016
- Levashina, J., & Campion, M. A. (2007). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, 14, 299-316. doi: 10.1111/j.1468-2389.2006.00353.x
- Levashina, J., Morgeson, F. P., & Campion, M. A. (2009). They don't do it often, but they do it well: Exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Assessment*, 17, 271-281. doi:10.1111/j.1468-2389.2009.00469.x

- Levin, R. A., & Zickar, M. J. (2002). Investigating self-presentation, lies, and bullshit: Understanding faking and its effects on selection decisions using theory, field research, and simulation. In J. M. Brett, & F. Drasgow (Eds.), *The psychology of work: Theoretically based empirical research* (pp. 253-276). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, *14*, 131-141. doi:10.1111/j.1468-2389.2006.00339.x
- Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology*, *110*, 401-412. doi:10.1037/0021-843X.110.3.401
- Mahar, D., Cologon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, *18*, 605-609. doi:10.1016/0191-8869(94)00200-C
- Maurer, T. J., Solamon, J. M., Andrews, K. D., & Troxtel, D. D. (2001). Interviewee coaching, preparation strategies, and response strategies in relation to performance in situational employment interviews: An extension of Maurer, Solamon, and Troxtel (1998). *Journal of Applied Psychology*, *86*, 709-717. doi:10.1037/0021-9010.86.4.709

- Maurer, T., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology*, 83, 128-136. doi:10.1037/0021-9010.83.1.128
- McCrae, R. R. (1986). Well-being scales do not measure social desirability. *Journal of Gerontology*, 41, 390-392.
- McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood*. New York, NY, US: Guilford Press.
- McCrae, R. R., & Costa, P. T. (1989). More reasons to adopt the five-factor model. *American Psychologist*, 44, 451-452. doi:10.1037/0003-066X.44.2.451
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90. doi:10.1037/0022-3514.52.1.81
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888. doi:10.1037/0022-006X.51.6.882
- McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). The uniform guidelines are a detriment to the field of personnel selection. *Industrial and Organizational Psychology*, 4, 494-514.
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment*, 11, 265-276. doi:10.1111/j.0965-075X.2003.00250.x

- McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*, 979-1016. doi:10.1111/j.0021-9029.2006.00052.x
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812-821. doi:10.1037/0021-9010.85.5.812
- McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment, 78*, 348-369. doi:10.1207/S15327752JPA7802_09
- McFarland, L. A., Yen, G., Harold, C. M., Viera, L., & Moore, L. G. (2005). An examination of impression management use and effectiveness across assessment center exercises: The role of competency demands. *Personnel Psychology, 58*, 949-980. doi:10.1111/j.1744-6570.2005.00374.x
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*, 531-552. doi:10.1348/0963179042596504
- Miller, C. E., & Barrett, G. V. (2008). The coachability and fakability of personality-based selection tests used for police selection. *Public Personnel Management, 37*, 339-351.
- Mischel, W. (1977). On the future of personality measurement. *American Psychologist, 32*, 246-254. doi:10.1037/0003-066X.32.4.246

- Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7, 109-117.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683-729. doi:10.1111/j.1744-6570.2007.00089.x
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029-1049. doi:10.1111/j.1744-6570.2007.00100.x
- Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Mumford, M. D., & Owens, W. A. (1982). Life history and vocational interests. *Journal of Vocational Behavior*, 21, 330-348. doi:10.1016/0001-8791(82)90041-0
- Murphy, K. R. (2009). Content validation is useful for many things, but validity isn't one of them. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 2, 453-464.
- Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than "g". In K. R. Murphy (Ed.), *Individual differences and behavior in organizations*. (pp. 3-30). San Francisco, CA: Jossey-Bass.

- Murphy, K. R., & Dzieweczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance, 18*, 343-357. doi:10.1207/s15327043hup1804_2
- Murphy, K. R., Dzieweczynski, J. L., & Yang, Z. (2009). Positive manifold limits the relevance of content-matching strategies for validating selection test batteries. *Journal of Applied Psychology, 94*, 1018-1031. doi: 10.1037/a0014075
- Nicholson, R. A., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist, 45*, 290-292. doi:10.1037/0003-066X.45.2.290
- O'Connell, V. (2009, January 7). Test for dwindling retail jobs spawns a culture of cheating. *The Wall Street Journal*, p. A1.
- Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology, 75*, 255-276. doi: 10.1348/096317902320369703
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995-1027. doi:10.1111/j.1744-6570.2007.00099.x
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609-626. doi: 10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679. doi:10.1037/0021-9010.81.6.660

- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-703.
doi:10.1037/0021-9010.78.4.679
- Oswald, F. L., & Hough, L. M. (2010). Personality and its assessment in organizations: Theoretical and empirical developments. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol. 2. Selecting and developing members for the organization* (pp. 153–184). Washington, DC: American Psychological Association.
- Oswald, F. L., & Hough, L. M. (2008). Personality testing and industrial–organizational psychology: A productive exchange and some future directions. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 323-332.
doi:10.1111/j.1754-9434.2008.00057.x
- Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64*, 569-607. doi:10.1037/0021-9010.64.5.569
- Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening applicants. *Personnel Psychology, 37*, 507-514. doi:10.1111/j.1744-6570.1984.tb00526.x
- Patrick, J. Characteristics of DSM-III borderline MMPI profiles. *Journal of Clinical Psychology, 40*, 655-658. doi: 10.1002/1097-4679(198405)40:3<655::AID-JCLP2270400304>3.0.CO;2-F

- Paulhus, D. L. (2002). *Socially desirable responding: The evolution of a construct*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Paulhus, D. L. (1991). *Measurement and control of response bias*. San Diego, CA, US: Academic Press.
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner, & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 143–165). New York: Springer-Verlag.
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, 21, 100-108. doi:10.1177/0146167295212001
- Payne, F. D. (1974). Relationship between response stability and item endorsement, social desirability, and ambiguity in the MMPI and CPI. *Multivariate Behavioral Research*, 9, 127-148. doi: 10.1207/s15327906mbr0902_1
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65, 70-89. doi:10.1177/0013164404268672
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172. doi: 10.1111/j.1744-6570.2008.00109.x.
- Ree, M. J., & Carretta, T. R. (2002). G2K. *Human Performance*, 15, 3-23.

- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1, 86-89. doi:10.1111/1467-8721.ep10768746
- Robie, C., & Ryan, A. M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment*, 7, 157-169. doi:10.1111/1468-2389.00115
- Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment*, 60, 215-226. doi:10.1207/s15327752jpa6002_1
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644. doi:10.1037/0021-9010.83.4.634
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-560. doi: 10.1111/j.1744-6570.1994.tb01736.x
- Roth, P. L., Switzer, F. S. III, Van Iddekinge, C. H., & Oh, I. (2011). Toward better meta-analytic matrices: How input values can affect research conclusions in human resource management simulations. *Personnel Psychology*, 64, 899-935.
- Rump, E. E., & Court, J. (1971). The Eysenck Personality Inventory and social desirability response set with student and clinical groups. *British Journal of Social & Clinical Psychology*, 10, 42-54.
- Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. L. Cooper, & I. T. Robertson (Eds.), *International review*

of industrial and organizational psychology 1989 (pp. 145-183). Oxford, England: John Wiley & Sons.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112-118.

Salgado, J. (2002). The big five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117-125. doi: 10.1111/1468-2389.00198

Schmit, M. J., & Ryan, A. M. (1993). The big five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 966-974. doi:10.1037/0021-9010.78.6.966

Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629-637. doi:10.1037/0021-9010.77.5.629

Schmitt, N., & Kuncce, C. (2002). The effects of required elaboration of answers to biodata questions. *Personnel Psychology*, 55, 569-587. doi:10.1111/j.1744-6570.2002.tb00121.x

Schmitt, N., & Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, 91, 613-621. doi:10.1037/0021-9010.91.3.613

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A., & Shivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology*, 92, 165-179. doi:10.1037/0021-9010.92.1.165

- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437-454.
- Schneider, R. J., Hough, L. M., & Dunnette, M. D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, 17, 639-655. doi:10.1002/(SICI)1099-1379(199611)17:6<639::AID-JOB3828>3.0.CO;2-9
- Sheppard, D., Smith, G. T., & Rosenbaum, G. (1988). Use of MMPI subtypes in predicting completion of a residential alcoholism–treatment program. *Journal of Consulting and Clinical Psychology*, 56, 590-596. doi: 10.1037//0022-006X.56.4.590
- SHL Group. (2000). *Expert report: OPQ32 Version n*. Boulder, CO: Author.
- SHL PreVisor (2011). *Do you want to profile performance or predict it: A commentary on profiling approaches to pre–employment assessment* [White Paper]. Retrieved from http://www.shl.com/assets/resources/White%20Paper_Profiling%20Predicting.pdf.
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87, 211-219. doi:10.1037/0021-9010.87.2.211
- Stanush, P. L. (1996). *Factors that influence the susceptibility of self-report inventories to distortion: A meta-analytic investigation*. Unpublished doctoral dissertation. College Station, TX: Texas A&M University.

- Storm, J., & Graham, J. R. (2000). Detection of coached general malingering on the MMPI—2. *Psychological Assessment, 12*, 158-165. doi:10.1037/1040-3590.12.2.158
- Tam, A. P., Murphy, K. R., & Lyall, J. T. (2004). Can changes in differential dropout rates reduce adverse impact? A computer simulation study of a multi-wave selection system. *Personnel Psychology, 57*, 905-934.
- Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56*, 621-633. doi: 10.1111/j.1467-6494.1988.tb00905.x
- Tellegen, A. (1982). *Brief manual for the Differential Personality Questionnaire*. Minneapolis: University of Minnesota Press.
- Tenopyr, M. L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology, 73*, 749-751. doi:10.1037/0021-9010.73.4.749
- Tett, R. P. (1998). Is conscientiousness ALWAYS positively related to job performance? *The Industrial–Organizational Psychologist, 36* (1), 24–29.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500-517. doi:10.1037/0021-9010.88.3.500
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967-993. doi:10.1111/j.1744-6570.2007.00098.x
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883. doi:10.1037/0033-2909.133.5.859
- van Hooft, E. A. J. & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, 97, 301-316. doi: 10.1037/a0025711
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, 19, 175-199. doi:10.1207/s15327043hup1903_1
- Vasilopoulos, N. L., Cucina, J. M., & Hunter, A. E. (2007). Personality and training proficiency: Issues of bandwidth-fidelity and curvilinearity. *Journal of Occupational and Organizational Psychology*, 80, 109-131. doi:10.1348/096317906X102114
- Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90, 306-322. doi:10.1037/0021-9010.90.2.306
- Velicer, W. F., & Weiner, B. J. (1975). Effects of sophistication and faking sets on the Eysenck Personality Inventory. *Psychological Reports*, 37, 71-73.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210. doi:10.1177/00131649921969802
- Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings?

International Journal of Selection and Assessment, 9, 277-289. doi:10.1111/1468-2389.00180

Warr, P., Bartram, D., & Martin, T. (2005). Personality and sales performance: Situational variation and interactions between traits. *International Journal of Selection and Assessment*, 13, 87-91. doi: 10.1111/j.0965-075X.2005.00302.x

Waters, S.D., & Sackett, P.R. (2006, May). *On the possibility of using configural scoring to enhance prediction*. Paper presented at the 21st Annual Conference Society of Industrial and Organizational Psychology, Dallas, Tx.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogenous and diverse task groups. *Academy of Management Journal*, 36, 590-602.

Wetter, M. W., Baer, R. A., Berry, D. T. R., Smith, G. T., & Larsen, L. H. (1992). Sensitivity of MMPI-2 validity scales to random responding and malingering. *Psychological Assessment*, 4, 369-374. doi:10.1037/1040-3590.4.3.369

White, L. A., Young, M. C., & Rumsey, M. G. (2001). ABLE implementation issues and related research. In J. P. Campbell, & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 525-558). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Whyte, W. H. (1975). How to cheat on personality tests. In A. Liebers (Ed.), *How to pass employment tests* (pp. 96-100). New York: ARCO.

- Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. D. (1984). *Multidimensional description of child personality: A manual for the Personality Inventory for Children* (rev. by D. Lachar). Los Angeles: Western Psychological Services.
- Witt, L. A., Burke, L. A., Barrick, M. A., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology*, 87, 164-169. doi: 10.1037//0021-9010.87.1.164
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, 18, 459-482. doi:10.1002/cne.920180503
- Young, M. C. (2003). *Effects of retesting on a new measure of motivational attributes: Implications for response distortion, test validity, and operational use*. Paper presented at the 27th annual meeting of the International Public Management Association Assessment Council, Baltimore, MD.
- Zellars, K. L., Perrewé, P. L., Hochwarter, W. A., & Anderson, K. S. (2006). The interactive effects of positive affect and conscientiousness on strain. *Journal of Occupational Health Psychology*, 11, 281-289. doi:10.1037/1076-8998.11.3.281
- Zickar, M.J. (2000). Modeling faking on personality tests. In Ilgen, D., & Hulin, C. L. (Eds). *Computational modeling of behavioral processes in organizations* (pp. 95 - 108). Washington, DC: American Psychological Association.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84, 551-563. doi:10.1037/0021-9010.84.4.551

- Zickar, M.J. & Slaughter, J.E. (2002). Computational modeling. In S.G. Rogelberg (Ed). *Research Methods in Industrial-Organizational Psychology* (pp. 184-197). Malden, MA: Blackwell.
- Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology*, 61, 309-348. doi: 10.1111/j.1744-6570.2008.00115.x
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz. (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). New York, NY: Academic Press.